

The accelerated gap times model

ROBERT L. STRAWDERMAN

*Department of Biological Statistics and Computational Biology,
Cornell University, Ithaca, NY 14853-7801, U.S.A.
rls54@cornell.edu*

SUMMARY

This paper develops a new semiparametric model for the effect of covariates on the conditional intensity of a recurrent event counting process. The model is a transparent extension of the accelerated failure time model for univariate survival data. Estimation of the regression parameter is motivated by semiparametric efficiency considerations, leading to a direct extension of the weighted log-rank estimating functions studied in Tsiatis (1990). A novel method for estimating the regression parameter using these rank-based estimating functions is proposed. An Aalen-type estimator for the baseline intensity function is also obtained. Asymptotics are handled through a combination of martingale and empirical process methods, and finite sample properties are studied via simulation. Finally, the new model is applied to the bladder tumor data of Byar (1980).

Some key words: Censoring; Counting process; Cox regression; Log-rank statistic; Multiple events; Rank regression; Recurrent events; Survival data.

1 INTRODUCTION

Recurrent events are prevalent in longitudinal studies in biomedical and public health settings, particularly cancer, AIDS, cardiovascular and renal disease, and mental health. Some examples of relevant recurrent phenomena include malignant tumor or cancer recurrence, episodic remissions, inpatient hospitalization, episodic infections or illness, nonfatal strokes and myocardial infarctions, schizophrenia, depression and epileptic and other seizures. A major goal when dealing with recurrent phenomena is to develop insight into the structure of the distribution of the time between event occurrences, or “gap” times, and how this distribution depends on important predictor variables including, perhaps, the past history of the event process itself.

Existing inference procedures for recurrent event data include both marginal and intensity-based methods. Marginal methods typically focus on the cumulative rate or mean function, and do not condition on the full event history. Recent examples include Pepe and Cai (1993), Lawless and Nadeau (1995), Lin, Wei and Ying (1998, 2001), and Lin, Wei, Yang, and Ying (2000). Typically, such methods are robust to the subject-level correlation structure between successive events, and are therefore useful if primary interest lies in understanding how population-level characteristics influence the marginal event rate at a fixed point in time. However, comparatively strong independent censorship conditions are also needed in order to avoid bias; see Kalbfleisch and Prentice (2002,

§9.2). Marginal models for the mean of a counting process also lose their attractive interpretation with time-dependent covariates, especially those that are functions of the past event history.

Intensity models fully specify how the probability of subsequent recurrence depends on the past event history, and are therefore less robust to the subject-level correlation structure between successive events. For the same reason, they are useful in studying the local dynamics of the recurrent event process and in particular for predicting recurrence experience at the subject level. Intensity models are also valid under comparatively weak noninformative censoring assumptions. “Andersen-Gill” intensity models (Andersen and Gill, 1982) assume the intensity of subsequent recurrence is independent of the past event history, given other observable possibly time-dependent factors. Such models can impose strong Markov assumptions on the underlying event process. Conditional intensity models allow the intensity of the next event to be more heavily dependent on the past event history, and are a more natural choice for problems involving multiple time scales. Semi-Markov intensity models, such as those described in Andersen, Borgan, Gill, and Keiding (1993, Example X.1.8), can substantially relax the assumptions of the Andersen-Gill model.

To date, the literature on conditional intensity models for recurrent event data has focused on semi-Markov models, parameterizing the event intensity in terms of the backward recurrence, or sojourn, time. The prevalent regression models are derived from the “modulated renewal process” of Cox (1972b), an extension of the Cox regression model for univariate survival data (Cox, 1972a). Specific examples can be found in Prentice, Williams, and Peterson (1981); Oakes (1992); Chang and Hsiung (1994); Oakes and Cui (1994); and Chang and Wang (1999). Nonparametric estimation in related settings has been considered by Gill (1980, 1981), Wang and Chang (1999), and Peña, Strawderman, and Hollander (2001).

In this paper a new semiparametric conditional intensity model is developed. This model directly extends the semiparametric accelerated failure time (AFT) model for survival data to recurrent event data. In particular, covariates serve to accelerate or decelerate the time to each recurrence, providing the implied intensity model with a “quite direct physical interpretation” (DR Cox, as quoted in Reid 1994, p. 450). In the case of time-independent covariates, the resulting intensity model, hereafter referred to as the accelerated gap times (AGT) model, explicitly depends on the past event history through the backward recurrence time. The extension of this intensity model to

time-dependent covariates broadens its scope, allowing the intensity to depend on general functions of the past event history. For example, the intensity of the next recurrence can depend on both the number and timing of previous events. There are several ways in which one might extend the proposed intensity model to handle time-dependent covariates. The approach taken here parallels the extension of the AFT model proposed by Robins and Tsiatis (1992).

The remainder of this paper will proceed as follows. In Section 2, the AGT model is developed for the case of time-independent covariates. In Section 3, semiparametric efficiency considerations are used to motivate an appropriate class of weighted logrank-type estimating functions for the regression parameter and baseline intensity function. The resulting class of estimating functions for the regression parameter extends those of Tsiatis (1990) to the case where each subject can experience more than one event. Relevant asymptotic theory is provided in Section 4. Novel methods of parameter and variance estimation are obtained as a byproduct of these results. These methods are easily adapted to the AFT model, hence providing some new methods for analyzing survival data using censored regression models. Section 5 studies the behavior of the proposed estimators via simulation, and Section 6 applies the proposed methods to the well-known bladder cancer study of Byar (1980). Finally, Section 7 concludes with some comments, including the necessary extensions for dealing with time-dependent covariates.

2 THE MODEL

2.1 Motivation and Definition

Consider first a single subject with a p -dimensional covariate vector Z experiencing repeated events at times $0 < U_1 < U_2 < \dots$. Define j^{th} gap time as $T_j = U_j - U_{j-1}$, where $j \geq 1$ and $U_0 = 0$. Suppose that, given Z , $T_j e^{\theta' Z} \stackrel{d}{=} V_j$, where V_1, V_2, \dots are independent and identically distributed random variables with absolutely continuous distribution function F_0 . The covariates Z directly accelerate or decelerate the “baseline” gap times V_j , $j \geq 1$, extending the AFT model in an obvious, intuitively appealing manner.

Define $U_m = \sum_{j=1}^m T_j$, and let $N^v(t) = \max(m : U_m \leq t)$ for $t \geq 0$. Given Z , $N^v(t)$ is a renewal process generated by the unobserved “baseline renewal process” $N^{b,v}(t) = \max(m : V_m \leq t)$ through a Z -dependent time transformation; that is, $N^{b,v}(t) \stackrel{d}{=} N^v(t e^{-\theta' Z})$ for all $t \geq 0$. Define

$\mathcal{F}_s = \sigma\{N^\nu(u), u \leq s; Z\}$ to be the history of the observed data to time s . Then, for any $s, \delta > 0$,

$$\text{pr}\{N^\nu(s + \delta) - N^\nu(s) = 1 | \mathcal{F}_s\} = \frac{F_0\{e^{\theta'Z} R^\nu(s + h)\} - F_0\{e^{\theta'Z} R^\nu(s)\}}{1 - F_0\{e^{\theta'Z} R^\nu(s)\}}$$

for $R^\nu(s) = s - U_{N^\nu(s-)}$. Dividing both sides by δ and letting $\delta \rightarrow 0$ yields the (predictable) local intensity for $N^\nu(\cdot)$ in the absence of censoring. With $\lambda_0(\cdot)$ denoting the hazard function for $F_0(\cdot)$, the corresponding cumulative intensity is

$$\Lambda(t|Z) = \int_0^t \lambda_0\{e^{\theta'Z} R^\nu(s)\} e^{\theta'Z} ds. \quad (1)$$

Using an obvious extension of notation, assume now that there are n subjects, the i^{th} being observed over $[0, \tau_i]$. The observed data are $\{N_i^\dagger(u), Y_i^\dagger(u), Z_i, u \geq 0\}$, $i = 1 \dots n$, where $N_i^\dagger(u) = N_i^\nu(u \wedge \tau_i)$, $Y_i^\dagger(u) = I(\tau_i \geq u)$, and $a \wedge b = \min(a, b)$. Throughout, it is assumed:

(A1) $F_0(\cdot) = \int_0^\cdot f_0(s)ds$, where $f_0(\cdot)$ has continuous first and bounded second derivatives,

survivor function $S_0(\cdot) = 1 - F_0(\cdot)$, hazard function $\lambda_0(\cdot)$, and renewal function $\gamma_0(\cdot)$;

(A2) $\{N_i^\dagger(u), Y_i^\dagger(u), Z_i, u \geq 0\}$, $i = 1 \dots n$ are independent and identically distributed;

(A3) $\tau_i \stackrel{ind}{\sim} G(\cdot|Z_i)$ for some possibly degenerate distribution $G(\cdot|Z_i)$;

(A4) Censoring is noninformative;

(A5) $\|Z_i\| < \infty$ a.s. and $E[\{N_1^\nu(\tau_1)\}^{6+\epsilon}] < \infty$ for some $\epsilon > 0$ and $t \geq 0$.

Let \mathcal{G}_t denote the event and covariate histories of all subjects to time t . Lemma 1, which will be taken as the operational definition of the AGT model, encapsulates the major model assumptions.

Lemma 1. For $i = 1, \dots, n$ define the backward recurrence time $R_i(u) = u - U_{iN_i^\dagger(u-)}$ and set

$$A_i^\dagger(t) = \int_0^t \lambda_0\{e^{\theta'Z_i} R_i(u)\} e^{\theta'Z_i} Y_i^\dagger(u) du. \quad (2)$$

Then, under (A1)-(A5), $M_i^\dagger(\cdot) = N_i^\dagger(\cdot) - A_i^\dagger(\cdot)$, $i \geq 1$ are orthogonal square integrable mean zero martingales with respect to $\{\mathcal{G}_t, t \geq 0\}$.

REMARK: The condition $E\{N_1^\nu(\tau)\} < \infty$ is sufficient to ensure square integrability holds (cf. Peña *et al.* 2001). The stronger condition in (A5) is only needed for the asymptotic results of Section 4. Since the moment generating function of $N^\nu(t)$ exists for all t (e.g., Resnick, 1994, Thm. 3.3.1), the condition $E[\{N_1^\nu(\tau)\}^{6+\epsilon}] < \infty$ restricts the decay rate of the upper tail of $G(\cdot|Z)$ and is easily met, for example, if τ has bounded support or a moment generating function.

2.2 Related models

The AGT model is related to some other models proposed in the literature. Assume for simplicity a single subject and no censoring. Then, the semi-Markov modulated renewal process model of Cox (1972b) assumes that the cumulative intensity function of $N^v(t)$ equals $\int_0^t \lambda_0\{s - U_{N^v(s-)}\} e^{\theta'Z} ds$; see also Oakes and Cui (1994). Models derived from or related to the modulated renewal process model include Prentice, Williams, and Peterson (1981); Oakes (1992); Chang and Hsiung (1994); and Chang and Wang (1999). Evidently, the main difference between the AGT model of Lemma 1 and the modulated renewal process model is that $e^{\theta'Z}$ appears as an argument to $\lambda_0(\cdot)$. These two models are equivalent if $\lambda_0(u) = \lambda_0$ for $u \geq 0$, and in this case both reduce to the Andersen-Gill model. Otherwise, all three intensity models are distinct and, comparatively speaking, the presence of $e^{\theta'Z}$ in $\lambda_0(\cdot)$ complicates both estimation and asymptotics in the AGT model.

Lin *et al.* (1998) propose an extension of the AFT model to deal with recurrent event data. In the notation of this paper, their “marginal accelerated means” (MAM) model assumes $E\{N^v(t)|Z\} = \mu_0(te^{\beta'Z})$, where $\mu_0(\cdot)$ is an unspecified increasing function and β is an unknown regression parameter. The MAM model neither assumes the gap times T_j are independent nor have the same conditional distributions (i.e., given Z). However, there is an interesting correspondence between the MAM and AGT models in the case where both models are valid. In particular, assuming the AGT model holds, $E\{N^v(t)|Z\} = \gamma_0(te^{\theta'Z})$, where $\gamma_0(t) = E\{N^{b,v}(t)\}$ is by (A1) the baseline renewal function. Hence, under the AGT model, $\mu_0(\cdot) = \gamma_0(\cdot)$ and the interpretation of θ and β coincide. It follows that the estimating function for β proposed in Lin *et al.* (1998) can also be used for estimating θ , provided censoring is independent of the underlying event process given Z . The estimating function (17) derived below is motivated by efficiency considerations and yields a consistent estimator under weaker censoring assumptions. The AGT model is also more natural choice if the focus of the analysis is on the recurrence time distribution. For example, with T denoting a generic gap time, it is easy to estimate

$$\text{pr}(T > t|Z) = \exp\{-\Lambda_0(te^{\theta'Z})\} \quad (3)$$

and, as a result, obtain median recurrence times and related quantities.

3 SEMIPARAMETRIC ESTIMATION

If $\Lambda_0(\cdot) = \int_0^\cdot \lambda_0(u)du$ is parametrically specified, standard likelihood methods may be used for parameter estimation, the corresponding asymptotic theory being obtained from a straightforward extension of the martingale convergence theory in Borgan (1984). For $\Lambda_0(\cdot)$ left unspecified, the AGT model is semiparametric and both estimation and asymptotics for the p -dimensional Euclidean parameter θ and infinite dimensional parameter $\Lambda_0(\cdot)$ become more difficult. Tsiatis (1990) proposed to estimate the regression parameter in the semiparametric AFT model by “inverting” a class of weighted linear rank statistics. Ritov (1990) studied this same problem from a different perspective, establishing a direct correspondence between the estimating functions of Tsiatis and those based on semiparametric efficiency considerations for censored linear regression models. In Section 3.1, semiparametric efficiency considerations are used to derive a weighted logrank-type estimating function for θ and also an estimator for $\Lambda_0(\cdot)$ under the AGT model. The resulting estimating function for θ directly generalizes those considered in Tsiatis (1990) to the case where each subject can experience more than one event.

3.1 Estimation – theoretical considerations

To motivate an estimating function for θ , we first obtain the semiparametric efficient score. These computations shall require that we be able to move freely between “calendar time” and “gap time” stochastic integral representations of the same quantity (cf. Peña *et al.*, 2001). Lemmas 2 and 3 provide the essential tools; the former is an easy consequence of the argument used to prove Proposition 1 of Peña *et al.* (2000), and the latter is proved in Appendix A.

Lemma 2. *Let (A1)-(A5) hold. For $s, t \geq 0$, $i = 1 \dots n$, define*

$$Y_i(s, t|\theta) = \sum_{j=1}^{N_i^\dagger(s-)} I(T_{ij}e^{\theta'Z_i} \geq t) + I[e^{\theta'Z_i}\{\tau_i \wedge s - U_{iN_i^\dagger(s-)}\} \geq t]. \quad (4)$$

Then, for any bounded integrable $h(\cdot)$,

$$\int_0^s h\{R_i(v)e^{\theta'Z_i}\}\lambda_0\{R_i(v)e^{\theta'Z_i}\}Y_i^\dagger(v)e^{\theta'Z_i}dv = \int_0^\infty h(r)\lambda_0(r)Y_i(s, r|\theta)dr.$$

To gain insight into the importance of this result, consider $h(r) = I(r \leq t)$ for some fixed $t > 0$, and assume $M_i^\dagger(\cdot)$ is a \mathcal{G}_s -martingale, i.e., so that θ is the “true” parameter in (2). Then,

$h\{R_i(v)e^{\theta'Z_i}\} = I\{R_i(v)e^{\theta'Z_i} \leq t\}$ is a \mathcal{G}_v -predictable process and $\int_0^s I\{R_i(v)e^{\theta'Z_i} \leq t\}M_i^\dagger(dv)$ is a \mathcal{G}_s -martingale process. It then follows from Lemma 2 that

$$M_i(s, t|\theta) = \int_0^s I\{R_i(v)e^{\theta'Z_i} \leq t\}M_i^\dagger(dv) \equiv N_i(s, t|\theta) - \int_0^t Y_i(s, u|\theta)\lambda_0(u)du, \quad (5)$$

where

$$N_i(s, t|\theta) = \int_0^s I\{R_i(v)e^{\theta'Z_i} \leq t\}N_i^\dagger(dv) \equiv \sum_{j=1}^{N_i^\dagger(s-)} I(T_{ij}e^{\theta'Z_i} \leq t) \quad (6)$$

and $N_i(s, t|\theta)$ and $Y_i(s, u|\theta)$ respectively track events and “at risk” status on the (baseline) gap time scale. Then, considered as a stochastic process, $M_i(s, t|\theta)$ is a mean zero \mathcal{G} -martingale in $s \geq 0$ for each fixed t . Since $E\{M_i(s, t|\theta)\} = 0$ for $i = 1 \dots n$, (5) suggests an obvious estimator for $\Lambda_0(\cdot)$ that, as shown below, can also be justified via maximum likelihood. However, despite the familiar form of (5), there is no filtration making $M_i(s, t|\theta)$ a martingale process in $t \geq 0$ for fixed s ; see Peña *et al.* (2001) for related discussion. Nevertheless, Lemma 3 shows these observations have useful consequences in more general moment calculations.

Lemma 3. Define $N_i(t|\theta) = N_i(\infty, t|\theta)$, $Y_i(t|\theta) = Y_i(\infty, t|\theta)$, and $M_i(t|\theta) = M_i(\infty, t|\theta)$. Let $H_i(\cdot)$ be a \mathcal{G} -predictable process and suppose $M_i^\dagger(\cdot)$ be a \mathcal{G} -martingale process, where \mathcal{G} is defined as in Lemma 1. Then, for any $t \geq 0$,

$$\int_0^t H_i(u)M_i(du|\theta) \equiv \int_0^\infty I\{R_i(v)e^{\theta'Z_i} \leq t\}H_i\{R_i(v)e^{\theta'Z_i}\}M_i^\dagger(dv).$$

Hence, $E\{\int_0^t H_i(u)M_i(du|\theta)\} = 0$ and

$$\text{var}\left\{\int_0^t H_i(u)M_i(du|\theta)\right\} = \int_0^t E\{H_i^2(u)Y_i(u|\theta)\}\lambda_0(u)du.$$

We now turn to the problem of computing the efficient score. Under the AGT model of Lemma 1, we may use Jacod’s point process loglikelihood (cf. Andersen *et al.*, 1993, Cor. II.7.3) to obtain the score for θ assuming $\Lambda_0(\cdot)$ is known. Specifically,

$$S_Q(\theta) = \sum_{i=1}^n \int_0^\infty Z_i Q\{R_i(u)e^{\theta'Z_i}\}M_i^\dagger(du) \quad (7)$$

where $M_i^\dagger(\cdot)$ is defined as in Lemma 1 and, here and henceforth, $Q(u) = 1 + u\{\dot{\lambda}_0(u)/\lambda_0(u)\}$ with $\dot{\lambda}_0(u)$ denoting the derivative of $\lambda_0(u)$. The semiparametric efficient score for θ is obtained by

subtracting from $S_Q(\theta)$ its orthogonal projection onto the tangent space for $\Lambda_0(\cdot)$; see, for example, Bickel *et al.* (1993, §3.4) or van der Vaart (1998, §25.4).

To compute the tangent space for $\Lambda_0(\cdot)$, consider the parameterized family of sub-models $(\theta, \Lambda_0) \rightarrow (\theta + \alpha\phi, \Lambda_\alpha)$, where $\alpha \in \mathbb{R}$ and $\dim(\phi) = \dim(\theta)$. Proceeding informally, suppose $\Lambda_\alpha(t) = \int_0^t \lambda_0(u) \{1 + \frac{\alpha}{2} \eta(u)\}^2 du$, where $\eta(\cdot)$ is a bounded element of

$$\mathbb{H} = \left\{ \eta : E \left[\int_0^\infty \eta \{R_1(u) e^{\theta' Z_1}\} M_1^\dagger(du) \right] = 0 \text{ and } E \left[\left(\int_0^\infty \eta \{R_1(u) e^{\theta' Z_1}\} M_1^\dagger(du) \right)^2 \right] < \infty \right\}. \quad (8)$$

Lemma 3 shows (8) is equivalent to

$$\mathbb{H} = \left\{ \eta : E \left[\int_0^\infty \eta(u) M_1(du|\theta) \right] = 0 \text{ and } \int_0^\infty \eta^2(r) \lambda_0(r) E\{Y_1(r|\theta)\} dr < \infty \right\}.$$

Notice that $(\theta + \alpha\phi, \Lambda_\alpha)$ reduces to (θ, Λ_0) at $\alpha = 0$. Under this family of sub-models, the score operator for (θ, Λ_0) (e.g., van der Vaart, 1998, §25.5.1) is $\phi' S_Q(\theta) + B_{\theta, \Lambda_0}(\eta)$, where $S_Q(\theta)$ and $Q(\cdot)$ are defined as above and

$$B_{\theta, \Lambda_0}(\eta) = \sum_{i=1}^n \int_0^\infty \eta \{R_i(u) e^{\theta' Z_i}\} M_i^\dagger(du) \quad (9)$$

is the score operator for $\Lambda_0(\cdot)$. The semiparametric efficient score for $\theta^{p \times 1}$ is now obtained by

1. determining $\varphi_k(\cdot) \in \mathbb{H}$, $k = 1 \dots p$ such that $E[\{S_{Qk}(\theta) - B_{\theta, \Lambda_0}(\varphi_k)\} B_{\theta, \Lambda_0}(\eta)] = 0$ for all $\eta \in \mathbb{H}$, where $S_{Qk}(\theta)$ is the k^{th} element of the p -vector $S_Q(\theta)$.
2. Computing $S_{eff}(\theta) = S_Q(\theta) - \{B_{\theta, \Lambda_0}(\varphi_1), \dots, B_{\theta, \Lambda_0}(\varphi_p)\}'$.

Proposition 1 below summarizes the results and is proved in Appendix A. Here and subsequently, $v^{\otimes k}$, $k = 0, 1, 2$ respectively denotes $1^{p \times 1}$, v , and vv' for a $p \times 1$ vector v ; in addition, $0/0 = 0$.

Proposition 1. *Let (A1)-(A5) hold and suppose $\int_0^\infty Q^2(u) \lambda_0(u) du < \infty$. Then, $\varphi_k(u) = Q(u) \mathcal{E}_k(u)$ for $\mathcal{E}_k(u) = E\{Z_{1k} Y_1(u|\theta)\} / E\{Y_1(u|\theta)\}$, $k = 1 \dots p$, and the semiparametric efficient score for θ*

$$\begin{aligned} S_{eff}(\theta) &= \sum_{i=1}^n \int_0^\infty Q\{R_i(u) e^{\theta' Z_i}\} [Z_i - \mathcal{E}\{R_i(u) e^{\theta' Z_i}\}] M_i^\dagger(du) \\ &\equiv \sum_{i=1}^n \int_0^\infty Q(u) \{Z_i - \mathcal{E}(u)\} M_i(du|\theta), \end{aligned} \quad (10)$$

where $\mathcal{E}(u) = E\{Z_1 Y_1(u|\theta)\} / E\{Y_1(u|\theta)\}$. Assume $\Sigma_Q = n^{-1} E\{S_{\text{eff}}(\theta) S'_{\text{eff}}(\theta)\} > 0$, where

$$\Sigma_Q = \int_0^\infty E[\{Z_1 - \mathcal{E}(u)\}^{\otimes 2} Y_1(u|\theta)] Q^2(u) \lambda_0(u) du. \quad (11)$$

Then, the semiparametric information bound for estimating θ is Σ_Q^{-1} .

The score (10) depends on the unknown function $Q(\cdot)$, which is difficult to estimate. Nevertheless, (10) provides an excellent starting point for developing a class of weighted estimating functions for θ , a problem tackled in the next section.

The score operator for $\Lambda_0(\cdot)$ is given in (9). Using Lemma 2, one may write this as

$$B_{\theta, \Lambda_0}(\eta) = \sum_{i=1}^n \int_0^\infty \eta(u) \{N_i(du|\theta) - Y_i(u|\theta) \Lambda_0(du)\}.$$

where $\eta(\cdot) \in \mathbb{H}$ is bounded. Solving $B_{\theta, \Lambda_0}(\eta) = 0$ for $\Lambda_0(\cdot)$ is therefore equivalent to solving $\sum_{i=1}^n N_i(du|\theta) - \Lambda_0(du) \{\sum_{i=1}^n Y_i(u|\theta)\} = 0$, yielding the Aalen-type estimator

$$\hat{\Lambda}_0(t|\theta) = \int_0^t \frac{\sum_{i=1}^n N_i(du|\theta)}{\sum_{i=1}^n Y_i(u|\theta)}. \quad (12)$$

Suppose θ is known and that $s^{(0)}(u|\theta) = E\{Y_1(u|\theta)\} > 0$ for $u \in [0, t]$. Then, under suitable regularity, (12) is evidently an asymptotically efficient estimator of $\Lambda_0(t)$ with efficient score

$$\sum_{i=1}^n \int_0^t \frac{M_i(du|\theta)}{s^{(0)}(u|\theta)} \equiv \sum_{i=1}^n \int_0^\infty \frac{I\{R_i(u) e^{\theta' Z_i} \leq t\}}{s^{(0)}\{R_i(u) e^{\theta' Z_i} | \theta\}} M_i^\dagger(du), \quad (13)$$

the latter representation following from Lemma 3. Notably, (13) equals $B_{\theta, \Lambda_0}(\eta_{\text{eff}, t})$, where $\eta_{\text{eff}, t}(u) = I(u \leq t) / s^{(0)}(u|\theta)$ is an element of \mathbb{H} . However, if θ is unknown, a consistent estimator for θ must be substituted in for θ and $\hat{\Lambda}_0(t|\hat{\theta})$ is no longer efficient in general. Proposition 2 provides the information bound for estimating $\Lambda_0(t)$ in this case.

Proposition 2. *Let the conditions of Proposition 1 hold. Define $s^{(1)}(u|\theta) = E\{Z_1 Y_1(u|\theta)\}$ and suppose $s^{(0)}(u|\theta) > 0$ for $u \in [0, t]$. Then, the semiparametric information bound for $\Lambda_0(t)$ is*

$$\int_0^t \frac{\lambda_0(u)}{s^{(0)}(u|\theta_0)} du + A'_Q(t) \Sigma_Q^{-1} A_Q(t), \quad (14)$$

where

$$A_Q(t) = \int_0^t \frac{s^{(1)}(u|\theta)}{s^{(0)}(u|\theta)} Q(u) \lambda_0(u) du. \quad (15)$$

3.2 Estimation – practical considerations

In light of (12), one may simply estimate $\Lambda_0(t)$ using $\widehat{\Lambda}_0(t|\widehat{\theta})$, where $\widehat{\theta}$ is any consistent estimator of θ . The main challenge therefore lies in estimating θ . From (10), observe that estimating θ requires estimation $\mathcal{E}(\cdot)$ and $Q(\cdot)$. Estimation of $\mathcal{E}(\cdot)$ is straightforward. Specifically, define

$$\mathbb{S}^{(k)}(t|\theta) = n^{-1} \sum_{j=1}^n Z_j^{\otimes k} Y_j(t|\theta); \quad (16)$$

then, one may estimate $\mathcal{E}(t)$ using $\mathbb{E}(t|\theta) = \mathbb{S}^{(1)}(t|\theta)/\mathbb{S}^{(0)}(t|\theta)$. Nonparametric estimation of $Q(\cdot)$ is more difficult. Given a nonparametric estimator for $\Lambda_0(\cdot)$, estimation of $Q(\cdot)$ using kernel methods is a possibility. However, such estimators are known to suffer from bias, are sensitive to bandwidth selection, and exhibit slow rates of convergence (e.g, Scott, 1992, §6.1-6.2). Thus, $Q(\cdot)$ can be difficult to estimate well with sample sizes arising in practical applications.

In the case of the AFT model, Tsiatis (1990) and Ritov (1990) replace the unknown optimal weight function with a data-dependent real-valued weight $W(u|\theta)$ that is easier to compute. In the current setting, and in view of (10), doing so generates the weighted class of estimating functions

$$S_W(\theta) = \sum_{i=1}^n \int_0^\infty W(u|\theta) \{Z_i - \mathbb{E}(u|\theta)\} M_i(du|\theta) \equiv \sum_{i=1}^n \int_0^\infty W(u|\theta) \{Z_i - \mathbb{E}(u|\theta)\} N_i(du|\theta). \quad (17)$$

The second form is an easy consequence of the definitions $M_i(\cdot|\theta)$ and $\mathbb{E}(\cdot|\theta)$ given earlier. With $N_i = N_i^\dagger(\tau_i-)$ denoting the observed number of events for subject i , define $\tilde{X}_{ij}(\theta) = T_{ij}e^{\theta'Z_i}$ for $j \leq N_i$ and $\tilde{X}_{ij}(\theta) = (\tau_i - U_{iN_i})e^{\theta'Z_i}$ for $j = N_i + 1$. Then, using (6), we may write

$$S_W(\theta) = \sum_{i=1}^n \sum_{j=1}^{N_i} W(\tilde{X}_{ij}(\theta)|\theta) \left[Z_i - \frac{\sum_{k=1}^n \sum_{r=1}^{N_k+1} Z_k I\{\tilde{X}_{kr}(\theta) \geq \tilde{X}_{ij}(\theta)\}}{\sum_{k=1}^n \sum_{r=1}^{N_k+1} I\{\tilde{X}_{kr}(\theta) \geq \tilde{X}_{ij}(\theta)\}} \right]. \quad (18)$$

The correspondence between (17) and the estimating function of Tsiatis (1990) should now be self-evident. In comparison, Lin *et al.* (1998) employ

$$S_{LWY}(\theta) = \sum_{i=1}^n \sum_{j=1}^{N_i} \widetilde{W}(U_{ij}e^{\theta'Z_i}|\theta) \left\{ Z_i - \frac{\sum_{k=1}^n Z_k I(U_{ij}e^{\theta'Z_i} \leq \tau_k e^{\theta'Z_k})}{\sum_{k=1}^n I(U_{ij}e^{\theta'Z_i} \leq \tau_k e^{\theta'Z_k})} \right\}$$

where $U_{ij} = \sum_{l=1}^j T_{il}$ for $j \geq 1$ and $\widetilde{W}(\cdot|\theta)$ is some weight function. Evidently, $S_{LWY}(\theta)$ is distinct from (18) and, for an equivalent choice of weight, the results of Section 3.1 therefore suggest (18) should yield a more efficient estimator.

Because (18) is a rank-based estimating function, one typically defines $\hat{\theta}$ as a zero-crossing of $S_W(\theta)$ or minimizer of $\|S_W(\theta)\|$, where $\|v\| = (v'v)^{1/2}$ for a vector v . However, because $S_W(\theta)$ is not necessarily monotone, there may exist several distinct minimizers. In the case of the semiparametric AFT model, these problems are largely mitigated upon using the so-called ‘‘Gehan’’ weight; see, for example, Fyngenson and Ritov (1994). For the AGT model, this corresponds to selecting $W(u|\theta) \equiv S^{(0)}(u|\theta)$, and the optimization problem may then be put in correspondence with minimizing a convex objective function. This minimization is easily done using linear programming methods; see, for example, Lin *et al.* (1998). Hereafter, $\hat{\theta}_G$ will be used to denote the solution $S_W(\theta) = 0$ for $W(u|\theta) \equiv S^{(0)}(u|\theta)$. Use of the Gehan weight yields a numerically stable, consistent sequence of estimators. However, it does not necessarily yield an estimator with other desirable statistical properties. It is shown later how the former can successfully be used in pursuit of the latter.

Other interesting weight functions include $W(u|\theta) \equiv 1$ and $W(u|\theta) = \hat{S}_0^\rho(u)\{1 - \hat{S}_0(u)\}^\delta$, where $\hat{S}_0(\cdot)$ is a suitable estimator of $S_0(\cdot)$. The former produces a logrank-type estimating function, the latter corresponds to the $G^{\rho,\delta}$ -class of weighted logrank tests developed in Fleming and Harrington (1991, §7.2.1). Ritov (1990) established the optimality of the logrank estimating function for the AFT model $\log T = -\theta'Z + \epsilon$, where ϵ has an extreme value distribution. Identifying ϵ with $\log V$, the logrank estimating function is therefore optimal if $T = Ve^{-\theta'Z}$ and V has a Weibull distribution. This optimality extends directly to the AGT model. Specifically, consider $Q(u)$ appearing in (10), and let $\xi > 0$ be a finite constant. Solving $Q(u) = \xi$ for $\lambda_0(\cdot)$ (i.e., constant weight) is equivalent to solving the ordinary linear differential equation $\dot{\lambda}_0(u) + (1 - \xi)u^{-1}\lambda_0(u) = 0$. Assuming $\lambda_0(1) = \lambda^\xi\xi$ for any $\lambda > 0$, we obtain $\lambda_0(u) = \lambda\xi(\lambda u)^{\xi-1}$, or the hazard function for a Weibull distribution. Hence the logrank estimating function, hereafter denoted $S_{LR}(\theta)$, is optimal for a large and flexible class of baseline intensity functions.

The estimating function $S_{LR}(\theta)$ is not guaranteed to be component-wise monotone in θ . As a result, it may have multiple distinct solutions, some of which may be inconsistent (cf. Fyngenson and Ritov, 1994). In the next section, it is shown that the Gehan estimator $\hat{\theta}_G$ is consistent and asymptotically normal under very weak conditions. These results are then used to construct a $n^{\frac{1}{2}}$ -consistent sequence of solutions in the case of a general weight function.

4 ASYMPTOTIC THEORY

For the sake of initial discussion, define $\widehat{\theta}_W$ as (approximately) minimizing $\|\bar{S}_W(\bar{\tau}, \theta)\|$, where

$$\bar{S}_W(t, \theta) = n^{-1} \sum_{i=1}^n \int_0^t W(u|\theta) \{Z_i - \mathbb{E}(u|\theta)\} M_i(du|\theta). \quad (19)$$

and $\bar{\tau} > 0$ is possibly infinite. In the absence of further information, establishing the consistency of $\widehat{\theta}_W$ can require strong global conditions (Newey and McFadden, 1994, §2.1). For example, two convenient and standard assumptions include (1) $\theta \in \Theta_c$, where $\Theta_c \subset \mathbb{R}^p$ is a compact set known to contain the true parameter θ_0 ; and, (2) $\bar{S}_W(\bar{\tau}, \theta)$ converges uniformly in probability to a continuous limit for $\theta \in \Theta_c$. The first condition is obviously restrictive, and is often ignored in practice. In the case of the AGT model, the second requires $E\{Y_1(t|\theta)\}$ to be bounded away from zero for $(t, \theta) \in [0, \bar{\tau}] \times \Theta_c$. This ensures $\mathbb{E}(u|\theta)$ remains bounded, but could be a rather restrictive condition if either T_{ij} or τ_i has finite support and Θ_c has a large diameter.

Both compactness and uniform convergence can be relaxed considerably if the estimating function is the gradient of a convex objective function (e.g., Andersen and Gill, 1982, Appendix II; Newey and McFadden, 1994, §2.6) or, more generally, a monotone field (Ritov, 1991). The monotonicity of $\bar{S}_W(\bar{\tau}, \theta)$ for $W(u|\theta) = \mathbb{S}^{(0)}(u|\theta)$ is exploited below in order to establish consistency and weak convergence of $\widehat{\theta}_G$ under very weak conditions. Given a general weight function, a consistent and asymptotically normal “one-step” estimator for θ is then obtained. The advantages of using a one-step estimator are discussed after Theorem 2 below. A novel algorithm for computing the one-step estimator and its variance is then proposed. Finally, asymptotic theory for $\widehat{\Lambda}_0(\cdot|\widehat{\theta}_W)$ is obtained, along with a method for consistently estimating its variance. Several important auxiliary results, as well as proofs of the main theorems appearing below, can be found in Appendix B. Martingale convergence theory cannot be applied in this problem, and the asymptotics therefore utilize empirical process theory as developed in van der Vaart and Wellner (1996).

4.1 Inference for θ

Employing the Gehan-type weight $W(u|\theta) = \mathbb{S}^{(0)}(u|\theta)$, (17) becomes

$$\bar{S}_G(t, \theta) = \int_0^t \mathbb{S}^{(0)}(u|\theta) \left\{ n^{-1} \sum_{i=1}^n Z_i N_i(du|\theta) \right\} - \int_0^t \mathbb{S}^{(1)}(u|\theta) \left\{ n^{-1} \sum_{i=1}^n N_i(du|\theta) \right\}. \quad (20)$$

Letting $t \rightarrow \infty$, some algebra shows

$$\bar{S}_G(\infty, \theta) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^{N_i} \sum_{k=1}^n \sum_{r=1}^{N_k+1} (Z_i - Z_k) I\{\tilde{X}_{kr}(\theta) \geq \tilde{X}_{ij}(\theta)\},$$

recalling again the notation $N_i = N_i^\dagger(\tau_i -)$, $\tilde{X}_{ij}(\theta) = T_{ij}e^{\theta'Z_i}$ for $j \leq N_i$, and $\tilde{X}_{ij}(\theta) = (\tau_i - U_{iN_i})e^{\theta'Z_i}$ for $j = N_i + 1$ from Section 3.2. Evidently, $\bar{S}_G(\theta) = \bar{S}_G(\infty, \theta)$ is the gradient of the convex objective function (cf. Fyngenson and Ritov, 1994; Lin *et al.*, 1998)

$$L_G(\theta) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^{N_i} \sum_{k=1}^n \sum_{r=1}^{N_k+1} \{\log \tilde{X}_{kr}(\theta) - \log \tilde{X}_{ij}(\theta)\}^+ \quad (21)$$

for $a^+ = \max\{a, 0\} \equiv (|a| + a)/2$. The minimizers of $L_G(\theta)$ and $\|\bar{S}_G(\theta)\|$ are asymptotically equivalent. Moreover, the set of minimizers of $L_G(\theta)$ lies in a convex set of diameter $O(n^{-1})$ (Fyngenson and Ritov, 1994). Hence we focus on $\hat{\theta}_G = \operatorname{argmin} L_G(\theta)$, where the right-hand side is taken to be any value of θ that minimizes $L_G(\theta)$.

Let $\theta \in \Theta$, where Θ is any open convex subset of \mathbb{R}^p . Then, under (A1)-(A5), Corollary 1 of Appendix B and (20) imply $\bar{S}_G(\theta)$ converges uniformly in probability to

$$S_G(\theta) = \int_0^t E\{Y_1(u|\theta)\} dE\{Z_1 N_1(u|\theta)\} - \int_0^t E\{Z_1 Y_1(u|\theta)\} dE\{N_1(u|\theta)\}. \quad (22)$$

Let $\mathcal{N}(\theta_0)$ be some open neighborhood of θ_0 and define the $p \times p$ matrix $D_G(\theta) = (d/d\theta)S_G(\theta)$. Theorem 1 below establishes the consistency and weak convergence of $\hat{\theta}_G$ under the following additional regularity conditions:

(A6) $s^{(j)}(u|\theta) = E\{Z_1^{\otimes j} Y_1(u|\theta)\}$, $j = 0, 1, 2$ are continuous for $(u, \theta) \in \mathbb{R}^+ \times \mathcal{N}(\theta_0)$.

(A7) $D_G = D_G(\theta_0)$ is nonsingular.

Existence of D_G is of course implied in (A7); in fact, the proof of Theorem 1 below shows that (A1)-(A6) are enough to guarantee the existence and continuity of $D_G(\theta)$ at $\theta = \theta_0$.

Theorem 1. *Suppose (A1)-(A7) hold and assume $\theta_0 \in \Theta$, where Θ is any open convex subset of \mathbb{R}^p . Then, $\hat{\theta}_G = \operatorname{argmin} L_G(\theta)$ exists with probability tending to one, $\hat{\theta}_G \xrightarrow{p} \theta_0$, and $n^{\frac{1}{2}}(\hat{\theta}_G - \theta_0) \xrightarrow{d} N(0, \Gamma_G)$, where $\Gamma_G = (D_G^{-1})' \Sigma_G D_G^{-1}$,*

$$\Sigma_G = \int_0^\infty \left[\left\{ s^{(0)}(u|\theta_0) \right\}^2 s^{(2)}(u|\theta_0) - s^{(0)}(u|\theta_0) \left\{ s^{(1)}(u|\theta_0) \right\}^{\otimes 2} \right] \lambda_0(u) du, \quad (23)$$

and

$$D_G = - \int_0^\infty \left[s^{(0)}(u|\theta_0) s^{(2)}(u|\theta_0) - \left\{ s^{(1)}(u|\theta_0) \right\}^{\otimes 2} \right] Q(u) \lambda_0(u) du. \quad (24)$$

A method for consistently estimating Γ_G is provided in the next section. In contrast to much of the survival analysis literature, Theorem 1 does not impose a bounded (gap) time interval assumption. This useful consequence of selecting the Gehan weight has apparently gone unnoticed in the literature on rank regression for censored data. Such is not the case for a general weight function $W(\cdot|\theta)$, where it is further assumed

(A8) $\bar{\tau}$ satisfies $\inf\{t \in [0, \bar{\tau}] : s^{(0)}(t|\theta_0)\} > 0$.

(A9) For $(t, \theta) \in [0, \bar{\tau}] \times \mathcal{N}(\theta_0)$, $W(t|\theta)$ is of bounded variation and there exists a continuous, bounded deterministic function $w(t|\theta)$ such that $|W(t|\theta) - w(t|\theta)|$ converges to zero uniformly in probability.

The bounded gap time interval assumption (A8) may be interpreted as requiring that the expected “baseline number at risk” $s^{(0)}(t|\theta_0)$ is positive for $t \in [0, \bar{\tau}]$. By (33) of Lemma 4 in Appendix B, (A8) is satisfied if $S_0(\bar{\tau})\text{pr}(\tau_1 e^{\theta_0' Z_1} \geq \bar{\tau}) > 0$. Then, similarly to (22), (A1)-(A9) and Corollary 1 imply $\bar{S}_W(\theta) = \bar{S}_W(\bar{\tau}, \theta)$ converges uniformly in probability to a function $S_W(\theta)$ for $\theta \in \mathcal{N}(\theta_0)$; see (38). The proof of Theorem 2 below shows that (A1)-(A9) ensure $D_W(\theta) = (d/d\theta)S_W(\theta)$ exists and is continuous at $\theta = \theta_0$. Analogously to (A7), it is further required that

(A10) $D_W = D_W(\theta_0)$ is nonsingular.

The following one-step estimation result then holds under (A1)-(A10).

Theorem 2. *Suppose (A1)-(A10) hold and assume $\theta_0 \in \Theta$, where Θ is an open convex subset of \mathbb{R}^p . Let \hat{D}_W be any consistent estimator of D_W . Then,*

$$\hat{\theta}_W = \hat{\theta}_G - (\hat{D}_W^{-1})' \bar{S}_W(\hat{\theta}_G)$$

is $n^{\frac{1}{2}}$ -consistent and $n^{\frac{1}{2}}(\hat{\theta}_W - \theta_0) \xrightarrow{d} N(0, \Gamma_W)$, where $\Gamma_W = (D_W^{-1})' \Sigma_W D_W^{-1}$,

$$\Sigma_W = \int_0^{\bar{\tau}} w^2(u|\theta_0) v(u|\theta_0) s^{(0)}(u|\theta_0) \lambda_0(u) du, \quad (25)$$

$v(u|\theta_0) = s^{(2)}(u|\theta_0)/s^{(0)}(u|\theta_0) - \{s^{(1)}(u|\theta_0)/s^{(0)}(u|\theta_0)\}^{\otimes 2}$, and

$$D_W = - \int_0^{\bar{\tau}} w(u|\theta_0) v(u|\theta_0) s^{(0)}(u|\theta_0) Q(u) \lambda_0(u) du, \quad (26)$$

If $w(u|\theta_0) \equiv Q(u)$ and $\bar{\tau} \rightarrow \infty$ in Theorem 2, then both Σ_W and $-D_W$ reduce to (11), yielding an asymptotic version of the usual information matrix equality. In this case, $\Gamma_Q = \Sigma_Q^{-1}$ and $\hat{\theta}_Q$ is a semiparametric efficient estimator by Proposition 1.

There are at least two notable advantages of using the one-step estimator $\hat{\theta}_W$. The asymptotic results of Tsiatis (1990), Ritov (1990), Ying (1993), and Lin *et al.* (1998) all require θ_0 to lie in a known compact subset of \mathbb{R}^p . This assumption is not required for the one-step estimator of Theorem 2. The reasons for this are: (i) iteration starts with the $n^{\frac{1}{2}}$ -consistent estimator $\hat{\theta}_G$, which is eventually guaranteed to be in some small neighborhood of θ_0 ; and, (ii) no compactness assumption is required for $\hat{\theta}_G$ because (21) and its limit are both convex. The second advantage is purely computational: a simple convex minimization problem (i.e., obtaining $\hat{\theta}_G$) replaces the potentially ill-behaved problem of minimizing $\|\bar{S}_W(\bar{\tau}, \theta)\|$ directly.

A consistent estimator of D_W is required in order to compute $\hat{\theta}_W$. However, because this same estimator is also needed for estimating Γ_W , this can hardly be viewed as a serious computational disadvantage. Computation of $\hat{\theta}_W$ and its variance are dealt with in the next section.

4.2 Estimating Γ_G and Γ_W

The asymptotic variances in Theorems 1 and 2 depend on (23) and (25), which are easy to estimate, and the (asymptotic) derivatives (24) and (26). The dependence of D_G and D_W on $Q(\cdot)$, hence $\lambda_0(\cdot)$ and $\dot{\lambda}_0(\cdot)$, is what makes estimation of Γ_G and Γ_W challenging in practice.

Huang (2002) proposed an interesting method of variance estimation in a related rank-regression problem. The following algorithm (i.e. Steps 2-4) adapts this method to the problem of estimating Γ_G of Theorem 1.

ALGORITHM #1

1. Compute $\hat{\theta}_G$ by minimizing (21);

2. Estimate Σ_G via

$$\hat{\Sigma}_G = \int_0^\infty \left[\mathbb{S}^{(0)}(u|\hat{\theta}_G) \mathbb{S}^{(2)}(u|\hat{\theta}_G) - \left\{ \mathbb{S}^{(1)}(u|\hat{\theta}_G) \right\}^{\otimes 2} \right] \bar{N}^{(0)}(du|\hat{\theta}_G),$$

where $\bar{N}^{(0)}(\cdot|\theta)$ is defined in (32) of Appendix B;

3. Minimize $\|\bar{S}_G(\theta) - c_k\|$ to obtain $\tilde{\theta}_{Gk}$, $k = 1 \dots p$, where c_k denotes the k^{th} column of $\hat{\Sigma}_G^{1/2}$;

4. Compute $\hat{\Gamma}_G = \left(\hat{\Gamma}_G^{1/2} \right)^2$, where $\hat{\Gamma}_G^{1/2} = n^{1/2}(\tilde{\theta}_{G1} - \hat{\theta}_G, \dots, \tilde{\theta}_{Gp} - \hat{\theta}_G)$.

Huang (2002) refers to this method of variance estimation as “inverse numerical differentiation.” This method does not estimate D_G or its inverse directly; rather, it exploits the asymptotic linearity and continuity of $\bar{S}_G(\theta)$ (and its inverse) in shrinking neighborhoods of θ_0 . Under (A1)-(A5), the consistency of $\hat{\Sigma}_G$ is a consequence of Corollary 1. That $\hat{\Gamma}_G$ is a consistent estimator of Γ_G is a simple consequence of Theorem 1 and the arguments used to prove Theorem 4 of Huang (2002). The performance of $\hat{\Gamma}_G$ is evaluated in Section 5, where it is observed to provide a stable and accurate estimate of Γ_G for modest sample sizes.

Algorithm #1 requires $p+1$ minimizations of a perturbed non-smooth objective function. In the case of the Gehan weight, these minimization problems are well-behaved and can be implemented very efficiently. Algorithm #1 is also easily adapted to the case of a general weight function. However, most other choices of $W(\cdot|\theta)$ yield non-monotone estimating functions, and solving these perturbed minimization problems can be more problematic. For example, in addition to the lack of efficient general algorithms, the minimizers may exhibit unstable behavior. Below, an alternative method is proposed for approximating D_W directly. This method also relies on the asymptotic linearity of $\bar{S}_W(\theta)$ in shrinking neighborhoods of θ_0 . However, it avoids the aforementioned instability by making greater use of the properties of $\hat{\theta}_G$. Specifically, for any $0 < M < \infty$ and $\|\theta - \theta_0\| \leq Mn^{-1/2}$, (41) of Appendix B implies

$$\bar{S}_W(\theta) - \bar{S}_W(\theta_0) = -D'_W(\theta - \theta_0) + o_p(n^{-1/2}).$$

Conditionally on the data, let $\tilde{\theta} \sim N(\hat{\theta}_G, \hat{\Gamma}_G)$, where $k = 1 \dots B$ and the replications are independent. Then, applying the above identity twice, we (informally) have

$$\bar{S}_W(\tilde{\theta}_k) - \bar{S}_W(\hat{\theta}_G) = D'_W(\hat{\theta}_G - \tilde{\theta}_k) + \text{error}.$$

The above relation suggests treating D_W as the slope in a regression of the (multivariate) “response” $\bar{S}_W(\tilde{\theta}_k) - \bar{S}_W(\hat{\theta}_G)$, $k = 1 \dots B$ on the “covariates” $\hat{\theta}_G - \tilde{\theta}_k$, $k = 1 \dots B$. More precisely, D_W can be estimated with $\hat{D}_{W,B} = (\mathfrak{Z}'\mathfrak{Z})^{-1} \mathfrak{Z}'\mathfrak{Y}$, where

$$\mathfrak{Z} = \begin{bmatrix} 1 & (\hat{\theta}_G - \tilde{\theta}_1)' \\ 1 & (\hat{\theta}_G - \tilde{\theta}_2)' \\ \vdots & \vdots \\ 1 & (\hat{\theta}_G - \tilde{\theta}_B)' \end{bmatrix} \text{ and } \mathfrak{Y} = \begin{bmatrix} \{\bar{S}_W(\tilde{\theta}_1) - \bar{S}_W(\hat{\theta}_G)\}' \\ \vdots \\ \{\bar{S}_W(\tilde{\theta}_B) - \bar{S}_W(\hat{\theta}_G)\}' \end{bmatrix}$$

Theorem 3. *Let $\hat{D}_W = \lim_{B \rightarrow \infty} \hat{D}_{W,B}$. Then, under (A1)-(A10), $\hat{D}_W \xrightarrow{p} D_W$ and $\hat{\Gamma}_W = (\hat{D}_W^{-1})' \hat{\Sigma}_W \hat{D}_W^{-1}$ is a consistent estimator for $\Gamma_W = (D_W^{-1})' \Sigma_W D_W^{-1}$, where*

$$\hat{\Sigma}_W = \int_0^{\bar{\tau}} W(u|\hat{\theta}_W) \left[\frac{\mathbb{S}^{(2)}(u|\hat{\theta}_W)}{\mathbb{S}^{(0)}(u|\hat{\theta}_W)} - \left\{ \frac{\mathbb{S}^{(1)}(u|\hat{\theta}_W)}{\mathbb{S}^{(0)}(u|\hat{\theta}_W)} \right\}^{\otimes 2} \right] \bar{N}^{(0)}(du|\hat{\theta}_W).$$

The proof of this result may be found in Appendix B. For $W(u|\theta) = 1$ (i.e., the logrank weight), the estimator $\hat{\Gamma}_W$ is shown to work well in Section 5, where it is also compared with an appropriate version of Huang’s estimator. The main advantage of this approach compared to Huang’s method stems from the replacement of $p + 1$ potentially unstable optimizations with $p + 1$ convex optimizations (i.e., for computing $\hat{\theta}_G$ and $\hat{\Gamma}_G$) followed by simple least squares calculation. The main disadvantage is the additional computation required for computing $\bar{S}_W(\tilde{\theta}_k)$, $k = 1 \dots B$.

The main algorithm for computing $\hat{\theta}_W$ and $\hat{\Gamma}_W$ is now summarized:

ALGORITHM #2

1. Use algorithm #1 to obtain $\hat{\theta}_G$ and $\hat{\Gamma}_G$;
2. Compute $\hat{D}_{W,B}$ for sufficiently large B ;
3. Compute $\hat{\theta}_W = \hat{\theta}_G - \hat{D}_{W,B}^{-1} \bar{S}_W(\hat{\theta}_G)$;
4. Compute $\hat{\Gamma}_W$ as in Theorem 3, using $\hat{D}_{W,B}$ in place of \hat{D}_W .

4.3 Inference for Λ_0

The main result is obtained for the general weighted case; i.e., where $\Lambda_0(t)$ is estimated via $\hat{\Lambda}_0(\cdot|\hat{\theta}_W)$ and $\hat{\theta}_W$ satisfies Theorem 2. Conditions (A9) and (A10) can be omitted if $\hat{\theta}_W$ is replaced by $\hat{\theta}_G$.

Theorem 4. Suppose (A1)-(A10) hold and that $\hat{\theta}_W$ satisfies Theorem 2. Then, $n^{\frac{1}{2}}\{\hat{\Lambda}_0(\cdot|\hat{\theta}_W) - \Lambda_0(\cdot)\}$ converges weakly to a Gaussian process on $[0, \bar{\tau}]$ with covariance function

$$\sigma_{\Lambda_0}^2(s, t) = \int_0^{t \wedge s} \frac{\lambda_0(u)}{s^{(0)}(u|\theta_0)} du + A'_Q(t) \Gamma_W A_Q(s), \quad (27)$$

where $\Gamma_W = (D_W^{-1})' \Sigma_W D_W^{-1}$ is the asymptotic variance of $n^{\frac{1}{2}}(\hat{\theta}_W - \theta_0)$ and $A_Q(t)$ is defined in (15). Moreover, for any $0 < M < \infty$,

$$\sup_{\substack{t \in [0, \bar{\tau}] \\ n^{\frac{1}{2}} \|\theta - \theta_0\| \leq M}} \left| \hat{\Lambda}_0(t|\theta) - \hat{\Lambda}_0(t|\theta_0) + A'_Q(t)(\theta - \theta_0) \right| = o_p(n^{-1/2}). \quad (28)$$

If $w(u|\theta_0) = Q(u)$, Theorem 2 shows Γ_W reduces to Σ_Q^{-1} , the latter being defined in (11). In this case, Proposition 2 implies $\hat{\Lambda}_0(\cdot|\hat{\theta}_Q)$ is semiparametric efficient. More generally, $\Gamma_W \neq \Sigma_Q^{-1}$; hence, Proposition 2 and Theorem 4 imply that efficient estimation of $\Lambda_0(\cdot)$ requires that $\hat{\theta}_W$ be asymptotically equivalent to $\hat{\theta}_Q$.

The variance (27) for $s = t$ can be consistently estimated with

$$\hat{\sigma}_{\Lambda_0}^2(t) = \int_0^t \{S^{(0)}(u|\hat{\theta}_W)\}^{-2} \bar{N}^{(0)}(du|\hat{\theta}_W) + \hat{A}'(t) \hat{\Gamma}_W \hat{A}(t),$$

where $\hat{\Gamma}_W$ is estimated as in Theorem 3 and $\hat{A}(t)$ is any consistent estimator of $A_Q(t)$. The asymptotic linearity result (28) shows that $A_Q(\cdot)$ can be approximated in a manner similar to Theorem 3. Confidence intervals and bands for $\Lambda_0(t)$ can then be computed in standard fashion. For example, a $100(1 - \alpha)\%$ confidence interval for $\Lambda_0(t)$ is (cf. Andersen *et al.*, 1993, §IV.1.3)

$$\hat{\Lambda}_0(t|\hat{\theta}_W) \exp \left[\mp z_{\alpha/2} \hat{\sigma}_{\Lambda_0}(t) / \{n^{\frac{1}{2}} \hat{\Lambda}_0(t|\hat{\theta}_W)\} \right]$$

5 SIMULATION RESULTS

Two simulations were done to evaluate the performance of $\hat{\theta}_G$ and also $\hat{\theta}_W$ for the logrank weight $W(\cdot|\theta) = 1$, hereafter denoted $\hat{\theta}_{LR}$. In the first simulation, the gap times $T \sim Ve^{-0.5Z}$, where $V \sim \text{Exp}(1)$, $Z \sim \text{Bernoulli}(0.5)$, and $\tau \sim \text{Uniform}(0, 3.5)$. This setting is equivalent to the ‘‘Poisson process’’ simulation summarized in Table 1 of Lin *et al.* (1998). In the second simulation, $V \sim \text{Gamma}(0.75, 0.75)$ and $Z \sim \text{Normal}(0, 1)$; T is then computed as above, with $\tau \sim \text{Uniform}(0, 3.5)$. Sample sizes of $n = 50$ and $n = 100$ are considered, and 1000 replications are used. Evidently, $\theta_0 = 0.5$ in both cases; the true values of $\{\Lambda_0(1), \Lambda_0(2)\}$ are (1, 2) and (1.73, 3.19), respectively.

Table 1: Simulation Results

	$W(\cdot \theta) = 1$ (Logrank)				$W(\cdot \theta) = \mathbb{S}^{(0)}(\cdot \theta)$ (Gehan)			
	Exponential		Gamma		Exponential		Gamma	
	n=50	n = 100	n=50	n = 100	n=50	n = 100	n=50	n = 100
$ \text{Bias}(\hat{\theta}) $	0.004	0.001	0.000	0.002	0.007	0.006	0.003	0.004
$\text{se}(\hat{\theta})$	0.209	0.144	0.095	0.065	0.237	0.160	0.113	0.076
$E\{\widehat{\text{se}}(\hat{\theta})\}$ (Huang)	0.192	0.135	0.092	0.063	0.227	0.158	0.110	0.076
$E\{\widehat{\text{se}}(\hat{\theta})\}$ (Thm. 3)	0.209	0.141	0.097	0.065	—	—	—	—
Emp. Cov. (95%)	0.944	0.943	0.957	0.955	0.945	0.949	0.944	0.950
$ \text{Bias}\{\hat{\Lambda}_0(1)\} $	0.000	0.006	0.001	0.004	0.006	0.000	0.002	0.001
$\text{se}\{\hat{\Lambda}_0(1)\}$	0.172	0.118	0.189	0.130	0.183	0.127	0.190	0.134
$E\{\widehat{\text{se}}\{\hat{\Lambda}_0(1)\}\}$	0.168	0.117	0.181	0.126	0.177	0.126	0.184	0.129
Emp. Cov. (95%)	0.938	0.958	0.953	0.947	0.933	0.957	0.950	0.944
$ \text{Bias}\{\hat{\Lambda}_0(2)\} $	0.008	0.000	0.003	0.003	0.003	0.003	0.034	0.015
$\text{se}\{\hat{\Lambda}_0(2)\}$	0.388	0.271	0.532	0.390	0.394	0.276	0.514	0.388
$E\{\widehat{\text{se}}\{\hat{\Lambda}_0(2)\}\}$	0.375	0.263	0.523	0.368	0.381	0.274	0.520	0.372
Emp. Cov. (95%)	0.950	0.948	0.944	0.947	0.952	0.959	0.948	0.942
$E\{Y_1(1 \theta_0)\}$	0.81				0.51			
$E\{Y_1(2 \theta_0)\}$	0.18				0.07			
$E\{N_1^\dagger(\tau_1)\}$	2.3				3.7			

The simulations were done in MATLAB 6; the results are summarized in Table 1. Respectively, the results for the Logrank and Gehan weights for a given combination of baseline distribution and sample size (i.e., columns (2,6), (3,7), (4,8), and (5,9)) are based on the same 1000 datasets. The estimators $\hat{\theta}_G$ and $\hat{\Gamma}_G$ are obtained using Algorithm #1, and the empirical bias and standard error are reported in the table. The estimators $\hat{\theta}_{LR}$ and $\hat{\Gamma}_{LR}$ are obtained using Algorithm #2. For comparison, a version of Algorithm #1 (i.e., steps 2-4) appropriate for the case of the logrank weight is also used to approximate the variance of $\hat{\theta}_{LR}$; the empirical mean of the Huang-type estimators for $\hat{\theta}_G$ and $\hat{\theta}_{LR}$ are given in the line labeled “ $E\{\widehat{\text{se}}(\hat{\theta})\}$ (Huang).” The entries for “ $E\{\widehat{\text{se}}(\hat{\theta})\}$ (Thm. 3)” refer to standard error approximations based on Theorem 3; these entries are blank for the Gehan weight because no such estimator was proposed for this case. In estimating the standard deviation of $\hat{\theta}_{LR}$, $B = 100$ simulated values are used per dataset; for $\hat{\Lambda}(\cdot|\hat{\theta}_G)$ and $\hat{\Lambda}(\cdot|\hat{\theta}_{LR})$, $B = 250$ simulated values are used. Also reported in the table are the average “expected baseline number at risk” at times $t = 1, 2$ and observed events per subject.

The proposed estimation methods work quite well for practical sample sizes. In particular, all estimators are nearly unbiased and confidence interval coverage agrees nicely with the nominally specified level of 95%. In the case of the logrank weight, there is no evidence that utilizing all of the data creates bias, suggesting that one may let $\bar{\tau} \rightarrow \infty$ in Theorem 2. The standard error of $\hat{\theta}_{LR}$ decreases when moving from Exponential to Gamma gap times for a given sample size; this is to be expected because there is an increase in the expected number of events per subject. In contrast, the variance of the cumulative hazard estimates increase, reflecting the reduced expected number at risk in the case of Gamma-distributed gap times at times $t = 1, 2$. Despite the use of simulation in approximating the asymptotic variance Γ_{LR} via Theorem 3, the standard error of this variance estimator never exceeded that of Huang’s estimator (results not shown).

In both simulations, the interpretation and numerical value of the true regression parameter coincide for the AGT and MAM models. A direct comparison may thus be made between the first 9 rows of Table 1 (columns labeled Exponential) and Table 1 of Lin *et al.* (1998) (columns labeled “Poisson Process”). In fact, the asymptotic variances can be computed exactly in this setting. Using the (optimal) logrank weight function, $\Gamma_{LR} \doteq 1/0.539$. For $n = 50$, we obtain an approximate standard error of $1.36(50)^{-1/2} \approx 0.192$ for $\hat{\theta}_{LR}$, very close to that reported in Table 1. For the estimator of Lin *et al.* (1998), say $\tilde{\theta}_{LR}$, we obtain an asymptotic variance and standard error of approximately $1/0.491$ and $1.43(50)^{-1/2} \approx 0.202$. Hence, $\hat{\theta}_{LR}$ yields an efficiency improvement of approximately 10% over $\tilde{\theta}_{LR}$ in this example. In the case of Gamma(0.75,0.75)-distributed baseline gap times, analytical computation of the asymptotic variances is not possible and $\tilde{\theta}_{LR}$ is compared to $\hat{\theta}_{LR}$ via simulation for $n = 50$. Using the (nonoptimal) logrank weight function, $\text{se}(\tilde{\theta}_{LR}) = 0.105$ and $\text{se}(\hat{\theta}) = 0.095$, suggesting an efficiency loss of 22%. Hence, in the two examples under consideration, estimation of θ_0 under the AGT model is more efficient than under the MAM model.

6 BLADDER TUMOR DATA ANALYSIS

In this section we illustrate the proposed methods using the bladder tumor cancer data of Byar (1980). The data, which come from a randomized clinical trial conducted by the Veterans Administration Co-operative Urological Group from 1971-1976, consist of 118 patients with bladder tumors. These tumors were removed, and patients were then randomized to one of three treatments: placebo, pyridoxine, or thiotepa. The data on thiotepa (38 patients, 45 total recurrences)

versus placebo (48 patients, 87 total recurrences) are analyzed in Lin *et al.* (1998), where they evaluate the effect of the baseline variables treatment (1 = placebo, 0 = thiotepa), number of initial tumors removed (integer, range 1-8), and diameter of largest initial tumor (in centimeters, range 1-8) on the average number of recurrences. Kalbfleisch and Prentice (2002, §9.4.3 and 9.5.2) also analyze these same data using several intensity and marginal models. Their analyses suggest that the risk of censoring is elevated for those patients experiencing a “recent recurrence,” i.e., tumor recurrence within the past month. Their analyses further suggest that the recurrence rate depends on the backward recurrence time. The latter conclusion is based on a Cox-type marginal model that includes these two functions of the past event history as time-dependent covariates, the covariate “recent recurrence” being required to adjust for the dependence of censoring on this variable. The importance of adjusting for recent recurrence is less clear in an intensity model utilizing backward recurrence time as the basic time scale. Hence, the data were reanalyzed under the AGT model using only baseline covariates. Table 2 contains the results, along with the estimates reported in Lin *et al.* (1998). The AGT model estimates are computed as described in Algorithms #1 and #2 of Section 4.2, using $B = 1500$ simulated values in Step 2 of Algorithm #2 for estimating \hat{D}_{LR} .

Table 2: Bladder Tumor Data

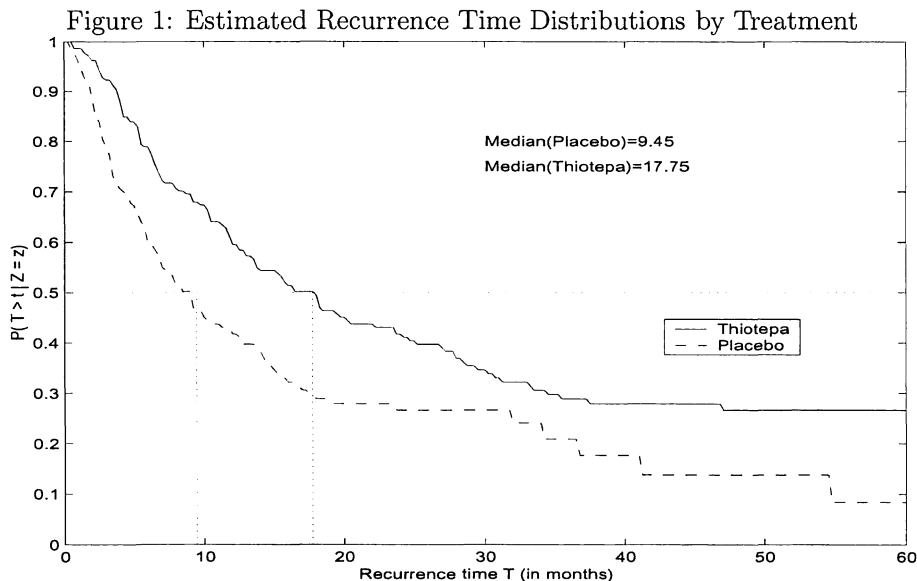
Weight	Covariate	AGT Model		MAM Model	
		$\hat{\theta}$	Est SE	$\hat{\theta}$	Est SE
Logrank	Treatment	0.681 [†]	0.278	0.542	0.312
	Initial #	0.264 [†]	0.067	0.204 [†]	0.066
	Initial Size	-0.031	0.096	-0.038	0.084
Gehan	Treatment	0.433	0.257	0.657	0.314
	Initial #	0.207 [†]	0.064	0.218 [†]	0.086
	Initial Size	-0.008	0.090	-0.022	0.101

[†] Significant at 5% level

The AGT and MAM models convey similar information, indicating tumor recurrence times expand on treatment (0 = thiotepa) and contract as the number of initial tumors increase. The treatment effect and initial number of tumors are both statistically significant at the 5% level under the AGT

model. The estimated standard errors under the two models are quite similar, though appreciably smaller for the treatment effect under the AGT model.

The AGT model provides the ability to estimate quantities directly related to the underlying recurrence time distribution; see, for example, (3). Figure 6 provides an estimate of (3) by treatment arm for the “average” patient (i.e., 2.5 initial tumors, the largest being 1.94 centimeters) using the logrank weight; also given are the estimated median times to tumor recurrence. The ordering observed may be partly artifact because the estimated distribution for the placebo group is obtained from that for the treatment group by a monotone time transformation. The estimated median time to tumor recurrence for the treated group is nearly double that of the placebo group, reflecting the effectiveness of thiotepa at reducing tumor recurrence. This is also indicated by the regression coefficient for the treatment effect, which suggests recurrence times for patients on thiotepa should be $\exp(0.681) \approx 2$ times as long as for patients on placebo.



For a hypothetical patient with fixed covariates \tilde{z} , it is possible to obtain a confidence interval for the median time to recurrence. Let $\tilde{\theta}_W$ and $\tilde{\Lambda}_0(t|\tilde{\theta}_W)$ denote the regression parameter and associated cumulative baseline intensity estimates based on the data $\{N_i^\dagger(u), Y_i^\dagger(u), \tilde{Z}_i, u \geq 0\}$, $i = 1 \dots n$, where $\tilde{Z}_i = Z_i - \tilde{z}$. Under this covariate translation $\Lambda_0(t)$ corresponds to the cumulative intensity for a patient having covariates \tilde{z} ; see also Lin *et al.* (1998), who point out this correspondence in the context of the MAM model. A family of 95% confidence intervals for the median recurrence

time of a patient with covariates \tilde{z} is then (cf. Andersen *et al.*, 1993, §IV.3)

$$\left\{ \xi \in \mathbb{R} : \frac{\left| g\{\tilde{\Lambda}_0(\xi|\tilde{\theta}_W)\} - g(\log 2) \right|}{\left| \dot{g}\{\tilde{\Lambda}_0(\xi|\tilde{\theta}_W)\} \right| \tilde{\sigma}_n(\xi)} \leq 1.96 \right\},$$

where $g(\cdot)$ is any suitable monotone transformation and $\tilde{\sigma}_n(\cdot)$ is the estimated standard deviation of $\tilde{\Lambda}_0(\cdot|\tilde{\theta}_W)$. For example, with $g(x) = \log(x)$ the desired interval is simply

$$\left\{ \xi \in \mathbb{R} : \tilde{\Lambda}_0(\xi|\tilde{\theta}_W) - \frac{1.96 \tilde{\sigma}_n(\xi)}{\tilde{\Lambda}_0(\xi|\tilde{\theta}_W)} \leq \log \log 2 \leq \tilde{\Lambda}_0(\xi|\tilde{\theta}_W) + \frac{1.96 \tilde{\sigma}_n(\xi)}{\tilde{\Lambda}_0(\xi|\tilde{\theta}_W)} \right\}. \quad (29)$$

Using (29), 95% confidence intervals for the median time to recurrence are (10.5, 27.0) and (6.5, 12.5) months for an average patient respectively treated with thiotepa or placebo.

7 DISCUSSION

The accelerated gap times model is a new intensity model for recurrent event data. It is a transparent extension of the semiparametric AFT model, and parallels the modulated renewal process generalization of the Cox regression model for univariate survival data. The intensity-based approach represents an alternative and useful perspective on efficient estimation in AFT-type models, culminating in a natural extension of the class of weighted estimating functions for θ considered by Tsiatis (1990) and Ritov (1990).

The idea to employ a one-step estimator is borrowed from Bickel *et al.* (1993, Ex. 7.7.2), who suggest using this method in the censored linear regression problem of Ritov (1990). However, they fail to point out the theoretical and practical advantages of doing so in this particular class of problems and, importantly, provide no guidance on how one might estimate the required analog of D_W . The method of Section 4.2 for estimating D_W in the case of a general weight function is new. Though it ostensibly assumes $\hat{\theta}_G \sim N(0, \hat{\Gamma}_G)$, the proposed method works with an arbitrary $n^{\frac{1}{2}}$ -consistent estimator $\hat{\theta}$ and does not require that one simulate draws from a normal distribution. The utility of this method is also not limited to the case of rank regression. For example, the method could be employed in a maximum likelihood estimation problem as a way to estimate the information matrix, and also in more general M -estimation problems as a way to estimate the “bread” in a sandwich estimator of variance.

In the case of time-independent covariates, the AGT model depends on the past history of the event process through the backward recurrence time only. The extension of this model to time-dependent

covariates allows the intensity to depend on general functions of the past event history and also further relaxes the requisite assumptions on censoring. There are several ways to “extend” the AGT model to the case of time-dependent covariates. For example, one might simply substitute $Z_i(u)$ in for Z_i in (2). However, the resulting intensity model does not correspond to a natural extension of the model of Robins and Tsiatis (1992). To obtain the desired extension, consider again the case of a single subject and define $\bar{Z} = \bar{Z}(\infty)$ as the history of an external time-dependent covariate $Z(t)$, $t \geq 0$. Proceeding similarly to Section 2, define $T_i = S_i - S_{i-1}$ for $i \geq 1$ ($S_0 = 0$) and $V_i = \int_{S_{i-1}}^{S_i} \exp\{\theta' Z(u)\} du$. Given \bar{Z} , suppose $\{V_i, i \geq 1\}$ are independent random variables with distribution function F_0 . Then, with $N^\nu(t)$ as defined in Section 2 and $\mathcal{F}_t = \sigma\{N^\nu(u), u \leq t; \bar{Z}\}$,

$$\Pr\{N^\nu(t+h) - N^\nu(t) = 1 | \mathcal{F}_t\} = \lim_{\delta \downarrow 0} \delta^{-1} \frac{\Pr\{t < S_{N^\nu(t)} + T_{N^\nu(t)} \leq t + \delta | \mathcal{F}_t\}}{\Pr\{S_{N^\nu(t)} + T_{N^\nu(t)} > t | \mathcal{F}_t\}}.$$

Define $g(t, \bar{Z}) = \int_0^t \exp\{\theta' Z(u)\} du$. Then, given \mathcal{F}_t , $g(S_{N^\nu(t)}, \bar{Z})$ is simply a known location shift and thus $g(S_{N^\nu(t)} + T_{N^\nu(t)}, \bar{Z})$ has the same probability distribution as $V + g(S_{N^\nu(t)}, \bar{Z})$. Because $g(t, \bar{Z})$ is monotone increasing in t , it is now easy to show

$$\frac{\Pr\{t < S_{N^\nu(t)} + T_{N^\nu(t)} \leq t + \delta | \mathcal{F}_t\}}{\Pr\{S_{N^\nu(t)} + T_{N^\nu(t)} > t | \mathcal{F}_t\}} = \frac{F_0\{g(t + \delta, \bar{Z}) - g(S_{N^\nu(t)}, \bar{Z})\} - F_0\{g(t, \bar{Z}) - g(S_{N^\nu(t)}, \bar{Z})\}}{1 - F_0\{g(t, \bar{Z}) - g(S_{N^\nu(t)}, \bar{Z})\}}.$$

Dividing both sides by δ , taking the limit as $\delta \downarrow 0$, and integrating yields the cumulative intensity function (i.e., with respect to \mathcal{F}_t), or

$$\Lambda_Z(t) = \int_0^t \lambda_0\{g(u, \bar{Z}) - g(S_{N^\nu(u-)}, \bar{Z})\} e^{\theta' Z(u)} du. \quad (30)$$

Notice that (30) reduces to (1) if $Z(u) = Z$ for $u \geq 0$. Moreover, since $g(S_k, \bar{Z}) \equiv \sum_{j=1}^k V_j$ (i.e., given \bar{Z}), $N^\nu(t) = \max\{k : S_k \leq t\} \equiv \max\{k : \sum_{j=1}^k V_j \leq g(t, \bar{Z})\}$. Hence, $E[N^\nu(t) | \bar{Z}] = \gamma_0(g(t, \bar{Z}))$ and corresponds to the extension of the MAM model considered by Lin *et al.* (1998).

The general AGT intensity model under noninformative censoring asserts that $\Lambda_Z(t \wedge \tau)$ is the compensator of $N^\nu(t \wedge \tau)$ with respect to $\mathcal{F}_t^* = \sigma\{N^\nu(u \wedge \tau), I(\tau \geq u), \bar{Z}(u+), u \leq t\}$. In principle, estimation with internal and external time-dependent covariates proceeds similarly to the case of time-independent covariates. Using a likelihood-based derivation similar to Section 3.1, the general form of the estimating function for θ quickly becomes complicated for arbitrary predictable $Z(\cdot)$ and, in particular, does not appear to reduce to the most obvious extension of (17). However, in the case where $Z_i(\cdot)$ changes only at event times, the general estimating function for θ becomes

computationally identical to (18). Specifically, define $\tilde{X}_{ij}(\theta) = T_{ij} \exp(\theta' Z_{ij})$ and $Z_{ij} = Z_i(S_{ij})$ for $1 \leq j \leq N_i$ and $\tilde{X}_{ij}(\theta) = (\tau_i - S_{iN_i}) \exp(\theta' Z_{i,N_i+1})$ and $Z_{i,N_i+1} = Z_i(\tau_i)$ for $j = N_i + 1$. Then, for a general weight function, the estimating function for θ reduces to

$$S_W(\theta) = \sum_{i=1}^n \sum_{j=1}^{N_i} W(\tilde{X}_{ij}(\theta)|\theta) \left[Z_{ij} - \frac{\sum_{k=1}^n \sum_{r=1}^{N_k+1} Z_{kr} I\{\tilde{X}_{kr}(\theta) \geq \tilde{X}_{ij}(\theta)\}}{\sum_{k=1}^n \sum_{r=1}^{N_k+1} I\{\tilde{X}_{kr}(\theta) \geq \tilde{X}_{ij}(\theta)\}} \right],$$

which is exactly (18) with Z_i replaced by Z_{ij} . Similarly, one obtains

$$\hat{\Lambda}_0(t|\theta) = \sum_{i=1}^n \sum_{j=1}^{N_i} \frac{I\{\tilde{X}_{ij}(\theta) \leq t\}}{\sum_{k=1}^n \sum_{r=1}^{N_k+1} I\{\tilde{X}_{kr}(\theta) \geq \tilde{X}_{ij}(\theta)\}}.$$

Several possibilities exist for extension of the AGT model. For example, paralleling Prentice *et al.* (1981), a history-dependent stratification of the baseline intensity $\lambda_0(\cdot)$ in (2) and also θ might be considered. However, eliminating the baseline intensity using a partial likelihood argument does not appear to be possible, complicating estimation. Exchangeable rather than iid baseline gap times could be accommodated using a shared frailty model, as was done in Peña *et al.* (2001); alternatively, a random effect could be incorporated directly into the multiplicative factor that accelerates the baseline gap times (cf., Cox and Solomon, 2003, §5.4). Clustered recurrent event processes, as one might encounter with patients treated at the same hospital or by the same physician, could be handled in a likelihood framework by extending the correlated frailty model of Parner (1998). Alternatively, a GEE-type approach might be employed using (17) and a working independence assumption. Finally, it is often true that observation of a recurrent event process is terminated by death. Extensions to this sort of dependent censoring problem would be worthwhile. However, because identifiability issues can arise when the death rate differs among patients having similar recurrent event rates, joint modeling of these processes is advisable.

APPENDIX A: PROOFS FOR §3.1

PROOF OF LEMMA 3:

For any $t \geq 0$, let $\mathcal{I}_1(t) = \int_0^t H_i(u) M_i(du|\theta)$ and $\mathcal{I}_2(t) = \int_0^\infty I\{R_i(v)e^{\theta'Z_i} \leq t\} H_i(R_i(v)e^{\theta'Z_i}) M_i^\dagger(dv)$. The equality $\mathcal{I}_1(t) = \mathcal{I}_2(t)$ is a consequence of (6), Lemma 2, and the fact that $M_i^\dagger(u) = N_i^\dagger(u) - \int_0^u Y_i^\dagger(v) \lambda_0(R_i(v)e^{\theta'Z_i}) e^{\theta'Z_i} dv$. Since $I\{R_i(v)e^{\theta'Z_i} \leq t\} H_i(R_i(v)e^{\theta'Z_i})$ is \mathcal{G}_v -predictable process, $E[\mathcal{I}_1(t)] = E[\mathcal{I}_2(t)] = 0$. In addition,

$$\text{var}(\mathcal{I}_1(t)) = E \left[\int_0^\infty H_i^2(R_i(v)e^{\theta'Z_i}) I\{R_i(v)e^{\theta'Z_i} \leq t\} Y_i^\dagger(v) \lambda_0(R_i(v)e^{\theta'Z_i}) dv \right].$$

Applying Lemma 2 once more establishes the variance identity and completes the proof. \square

PROOF OF PROPOSITION 1

We wish to find $\varphi_k(\cdot) \in \mathbb{H}$, $k = 1 \dots p$ such that $E[(S_{Qk}(\theta) - B_{\theta, \Lambda_0}(\varphi_k)) B_{\theta, \Lambda_0}(\eta)] = 0$ for all $\eta \in \mathbb{H}$. It suffices to consider $k = 1$ and $\eta(\cdot)$ not identically zero. The quantities $S_{Q1}(\theta)$ and $B_{\theta, \Lambda_0}(\eta)$ are stochastic integrals, and depend on the \mathcal{G} -martingales $M_i^\dagger(\cdot)$, $i = 1 \dots n$ defined as in Lemma 1. Since $M_i^\dagger(\cdot)$, $i = 1 \dots n$ are orthogonal mean zero square-integrable martingales and $Q(u)$ is deterministic, a straightforward martingale covariance calculation shows

$$\begin{aligned} E[B_{\theta, \Lambda_0}(\eta) (S_{Q1}(\theta) - B_{\theta, \Lambda_0}(\varphi_1))] &\equiv \\ nE \left[\int_0^\infty \eta(R_1(u)e^{\theta'Z_1}) \left\{ Z_{11} Q(R_1(u)e^{\theta'Z_1}) - \varphi_1(R_1(u)e^{\theta'Z_1}) \right\} \lambda_0(R_1(u)e^{\theta'Z_1}) Y_1^\dagger(u) e^{\theta'Z_1} du \right] \end{aligned}$$

The integral term on the right-hand side is easily handled via Lemma 2; hence,

$$E[B_{\theta, \Lambda_0}(\eta) (S_{Q1}(\theta) - B_{\theta, \Lambda_0}(\varphi_1))] \equiv nE \left[\int_0^\infty \{Z_{11} Q(r) - \varphi_1(r)\} \eta(r) \lambda_0(r) Y_1(r|\theta) dr \right]$$

Hence, $E[B_{\theta, \Lambda_0}(\eta) (S_{Q1}(\theta) - B_{\theta, \Lambda_0}(\varphi_1))] = 0$ if and only if $E[Z_{11} Y_1(r|\theta)] Q(r) = \varphi_1(r) E[Y_1(r|\theta)]$ or, equivalently, that $\varphi_1(r) = Q(r) \mathcal{E}_1(r)$ for $\mathcal{E}_1(r) = E[Z_{11} Y_1(r|\theta)] / E[Y_1(r|\theta)]$. Since Z_1 is uniformly bounded by (A5) and $0/0=0$, we have $\mathcal{E}_1(r) = 0$ whenever $E[Y_1(r|\theta)] = 0$; hence $\varphi_1(\cdot) \in \mathbb{H}$. The efficient score is now obtained by subtracting $[B_{\theta, \Lambda_0}(\varphi_1), \dots, B_{\theta, \Lambda_0}(\varphi_p)]'$ from $S_Q(\theta)$; (10) and (11) then follow directly from Lemma 3. The information bound is then a direct consequence of Theorem 3.4.1 of Bickel *et al.* (1993). \square

PROOF OF PROPOSITION 2

As implied in van der Vaart (1998, §25.5) and also by Corollaries 5.5.2 and 5.5.3 of Bickel *et al.* (1993), the efficient score for $\Lambda_0(\cdot)$ equals

$$B_{\theta, \Lambda_0}(\eta_{\text{eff}, t}) - E[B_{\theta, \Lambda_0}(\eta_{\text{eff}, t}) S'_Q(\theta)] (n\Sigma_Q)^{-1} S_{\text{eff}}(\theta) \quad (31)$$

where $B_{\theta, \Lambda_0}(\eta_{\text{eff}, t})$ is given by the right-hand side of (13), $S_Q(\theta)$ is given in (7), $S_{\text{eff}}(\theta)$ is given in Proposition 1, and Σ_Q is defined in (11). A straightforward computation shows

$$E[B_{\theta, \Lambda_0}(\eta_{\text{eff}, t}) S_Q(\theta)] = \sum_{i=1}^n E \left[\int_0^\infty \frac{I\{R_i(u)e^{\theta'Z_i} \leq t\}}{s^{(0)}(R_i(u)e^{\theta'Z_i}|\theta)} Z_i Q(R_i(u)e^{\theta'Z_i}) Y_i^\dagger(u) \lambda_0(R_i(u)e^{\theta'Z_i}) e^{\theta'Z_i} du \right]$$

Using Lemma 2 and simplifying, we find $E[B_{\theta, \Lambda_0}(\eta_{\text{eff}, t}) S_Q(\theta)] = nA_Q(t)$. Hence, (31) reduces to

$$B_{\theta, \Lambda_0}^{(\text{eff})}(t) = \sum_{i=1}^n \int_0^\infty \left[\frac{I\{u \leq t\}}{s^{(0)}(u|\theta)} - A'_Q(t) \Sigma_Q^{-1} Q(u) (Z_i - \mathcal{E}(u)) \right] M_i(du|\theta).$$

By Lemma 3 and the fact that the summands are iid,

$$\text{var} \left(n^{-1/2} B_{\theta, \Lambda_0}^{(\text{eff})}(t) \right) = E \left[\int_0^\infty \left\{ \frac{I\{u \leq t\}}{s^{(0)}(u|\theta)} - A'_Q(t) \Sigma_Q^{-1} Q(u) (Z_1 - \mathcal{E}(u)) \right\}^2 Y_1(u|\theta) \lambda_0(u) du \right].$$

Expanding the square and simplifying then yields (14). \square

APPENDIX B: PROOFS AND AUXILIARY RESULTS FOR §4

We begin with several fundamental results. All probability and expectations are with respect to \mathbb{P}_0 , the underlying probability measure associated with the true parameters θ_0 and $\Lambda_0(\cdot)$. Throughout, $\mathbb{R}^+ = [0, \infty)$, and Θ is a open, convex subset of \mathbb{R}^p . Unless otherwise stated, all proofs for §4 and auxiliary results provided below are collected at the end.

Theorem 5. *Let $\theta \in \Theta$ be fixed and define for $t \in \mathbb{R}^+$*

$$\bar{N}^{(j)}(t|\theta) = \frac{1}{n} \sum_{i=1}^n Z_i^{\otimes j} N_i(t|\theta), \quad j = 0, 1, 2. \quad (32)$$

Then, considered as processes in $t \in \mathbb{R}^+$, (16) and (32) are \mathbb{P}_0 -Donsker under (A1)-(A5).

Theorem 6. *As processes in $(t, \theta) \in \mathbb{R}^+ \times \Theta$, (16) and (32) are \mathbb{P}_0 -Donsker under (A1)-(A5).*

The following is an immediate consequence of Theorems 5 and 6.

Corollary 1. *Under conditions (A1)-(A5) and for $j = 0, 1, 2$,*

$$\sup_{(t, \theta) \in \mathbb{R}^+ \times \Theta} \left\| \bar{N}^{(j)}(t|\theta) - E \left[Z_1^{\otimes j} N_1(t|\theta) \right] \right\| \xrightarrow{P} 0 \quad \text{and} \quad \sup_{(t, \theta) \in \mathbb{R}^+ \times \Theta} \left\| \bar{Y}^{(j)}(t|\theta) - E \left[Z_1^{\otimes j} Y_1(t|\theta) \right] \right\| \xrightarrow{P} 0.$$

These results continue to hold if the supremum is taken only over $t \in \mathbb{R}^+$, $\theta \in \Theta$ fixed.

Lemma 4 below is useful for computing the expectations appearing in Corollary 1. The first result is an easy consequence of (5), (6), and Lemma 3; the expectation (33) follows immediately from Proposition 2 of Peña *et al.* (2001). Proof is therefore omitted.

Lemma 4. *Let $\theta \in \mathbb{R}^p$ be arbitrary, and suppose θ_0 is the unique parameter value making $M_i^\dagger(\cdot)$ in Lemma 1 a \mathcal{G} -martingale. Then, $N_i(t|\theta) = N_i(te^{(\theta_0 - \theta)' Z_i}|\theta_0)$ and $Y_i(t|\theta) = Y_i(te^{(\theta_0 - \theta)' Z_i}|\theta_0)$ for $i = 1 \dots n$. Moreover, for $k = 0, 1, 2$,*

$$E \left[Z_1^{\otimes k} N_1(t|\theta) \right] = E \left[Z_1^{\otimes k} \int_0^{te^{(\theta_0 - \theta)' Z_1}} Y_1(u|\theta_0) \lambda_0(u) du, \right]$$

where, for $H_{\theta_0}(r|Z_1) = pr\{\tau_1 e^{\theta'_0 Z_1} \leq r|Z_1\}$,

$$E[Y_1(t|\theta_0)|Z_1] = S_0(t) \int_t^\infty (1 + \gamma_0(r - t)) H_{\theta_0}(dr|Z_1). \quad (33)$$

PROOF OF THEOREM 1:

From (22), (A6), and Lemma 4, it follows that $\bar{S}_G(\theta) \xrightarrow{P} \mathcal{S}_G(\theta)$, where

$$\mathcal{S}_G(\theta) = \int_0^\infty s^{(0)}(u|\theta) E \left[Z_1 Y_1(u|\theta) \lambda_0(u h_1(\theta)) h_1(\theta) \right] - s^{(1)}(u|\theta) E \left[Y_1(u|\theta) \lambda_0(u h_1(\theta)) h_1(\theta) \right] du$$

and $h_1(\theta) = \exp\{(\theta_0 - \theta)'Z_1\}$. Evidently, $h_1(\theta_0) = 1$ and thus $\mathcal{S}_G(\theta_0) = 0$; (A6) and (A7) therefore ensure θ_0 is unique. Corollary 1 and Theorem 5 further imply

$$\bar{S}_G(\theta_0) = \frac{1}{n} \sum_{i=1}^n \int_0^\infty \left[s^{(0)}(u|\theta_0)Z_i - s^{(1)}(u|\theta_0) \right] M_i(du|\theta_0) + o_p(n^{-1/2}).$$

The first term on the right-hand side is an average of iid mean zero random variables. Using Lemma 3, it can be shown that the variance of $\int_0^\infty [s^{(0)}(u|\theta_0)Z_i - s^{(1)}(u|\theta_0)] M_i(du|\theta_0)$ reduces to Σ_G as stated in the theorem. Moreover, by (A1) and (A6), $\|\Sigma_G\| < \infty$, and hence the first term on the right-hand side satisfies an ordinary central limit theorem.

Define $C_M = \{\alpha \in \mathbb{R}^p : \|\alpha\| \leq M\}$, where $0 < M < \infty$ is arbitrary. Since $\bar{S}_G(\theta)$ and $\mathcal{S}_G(\theta)$ are (component-wise) monotone functions of θ , the theorem follows from the above and Theorem 24.2 of Ritov (1991) provided that (24) holds and

$$n^{\frac{1}{2}} \left\{ \bar{S}_G(\theta_0 + n^{-1/2}\eta) - \bar{S}_G(\theta_0) - \mathcal{S}_G(\theta_0 + n^{-1/2}\eta) \right\} = o_p(1) \quad (34)$$

for any $\eta \in C_M$. The result (34) follows if, for any (positive) sequence $\delta_n \rightarrow 0$,

$$\sup_{\|\theta - \theta_0\| \leq \delta_n} n^{\frac{1}{2}} \left\{ \bar{S}_G(\theta) - \bar{S}_G(\theta_0) - \mathcal{S}_G(\theta) \right\} = o_p(1 + n^{\frac{1}{2}}\|\theta - \theta_0\|). \quad (35)$$

To show (35), write $\bar{S}_G(\theta) = (I) + (II)$, where

$$(I) = \frac{1}{n} \sum_{i=1}^n \int_0^\infty \left[\mathbb{S}^{(0)}(u|\theta)Z_i - \mathbb{S}^{(1)}(u|\theta) \right] [N_i(du|\theta) - Y_i(u|\theta)\lambda_0(uh_i(\theta))h_i(\theta)du]$$

and

$$(II) = \frac{1}{n} \sum_{i=1}^n \int_0^\infty \left[\mathbb{S}^{(0)}(u|\theta)Z_i - \mathbb{S}^{(1)}(u|\theta) \right] Y_i(u|\theta) [\lambda_0(uh_i(\theta))h_i(\theta) - \lambda_0(u)] du.$$

From Lemma 4, $N_i(t|\theta) = N_i(te^{(\theta_0 - \theta)'Z_i}|\theta_0)$ and $Y_i(t|\theta) = Y_i(te^{(\theta_0 - \theta)'Z_i}|\theta_0)$. Substituting these representations into (I) and simplifying, we see $(I) = \bar{S}_G(\theta_0)$. Hence, $\bar{S}_G(\theta) - \bar{S}_G(\theta_0)$ reduces to (II). Now, suppose that $\|\theta - \theta_0\| \leq \delta_n$, where $\delta_n \rightarrow 0$. (A1) then implies

$$\lambda_0(uh_i(\theta))h_i(\theta) = \lambda_0(u) - q_0(u)Z_i'(\theta - \theta_0) + O(\|\theta - \theta_0\|^2) \text{ a.s.}$$

where $q_0(u) = \lambda_0(u) + u\dot{\lambda}_0(u)$ and the constant in the error term can be bounded independently δ_n . Substituting this expansion into (II) and simplifying, we obtain

$$\bar{S}_G(\theta) - \bar{S}_G(\theta_0) = -(\theta - \theta_0)' \int_0^\infty \left(\mathbb{S}^{(0)}(u|\theta)\mathbb{S}^{(2)}(u|\theta) - [\mathbb{S}^{(1)}(u|\theta)]^{\otimes 2} \right) q_0(u)du + o_p(\|\theta - \theta_0\|). \quad (36)$$

Because $\mathcal{S}_G(\theta_0) = 0$, an essentially identical argument also shows

$$\mathcal{S}_G(\theta) = -(\theta - \theta_0)' \int_0^\infty \left(s^{(0)}(u|\theta)s^{(2)}(u|\theta) - [s^{(1)}(u|\theta)]^{\otimes 2} \right) q_0(u)du + o(\|\theta - \theta_0\|). \quad (37)$$

Subtracting (37) from (36) and multiplying through by $n^{\frac{1}{2}}$, Theorem 6 and the Delta Method yield

$$\sup_{\|\theta - \theta_0\| \leq \delta_n} n^{\frac{1}{2}} \left\{ \bar{S}_G(\theta) - \bar{S}_G(\theta_0) - \mathcal{S}_G(\theta) \right\} = O(\delta_n) \times O_p(1) + o_p(n^{\frac{1}{2}}\|\theta - \theta_0\|).$$

Since $\delta_n \rightarrow 0$, the first term is $o_p(1)$, establishing (35). Observing that (24) is a direct consequence of (37) and that $q_0(u) = Q(u)\lambda_0(u)$ then completes the proof. \square

PROOF OF THEOREM 2:

Under the stated conditions, $\bar{S}_W(\theta) = \bar{S}_W(\bar{\tau}, \theta)$ converges uniformly to (i.e., for $\theta \in \mathcal{N}(\theta_0)$)

$$\mathcal{S}_W(\theta) = \int_0^{\bar{\tau}} w(u|\theta) \left(E[Z_1 Y_1(u|\theta) \lambda_0(u h_1(\theta)) h_1(\theta)] - \frac{s^{(1)}(u|\theta)}{s^{(0)}(u|\theta)} E[Y_1(u|\theta) \lambda_0(u h_1(\theta)) h_1(\theta)] \right) du \quad (38)$$

where $h_1(\theta) = \exp\{(\theta_0 - \theta)' Z_1\}$. Because $h_1(\theta_0) = 1$, $\mathcal{S}_W(\theta_0) = 0$; (A6) and (A10) therefore ensure θ_0 is unique. By Corollary 1 and Theorem 5

$$\bar{S}_W(\theta_0) = \frac{1}{n} \sum_{i=1}^n \int_0^{\bar{\tau}} w(u|\theta_0) \left[Z_i - \frac{s^{(1)}(u|\theta_0)}{s^{(0)}(u|\theta_0)} \right] M_i(du|\theta_0) + o_p(n^{-1/2}). \quad (39)$$

The first term on the right-hand side is an average of iid mean zero random variables. Using Lemma 3, a straightforward computation shows that the variance of a single term equals Σ_W . Under the stated assumptions, $\|\Sigma_W\| < \infty$ and the first term on the right-hand side satisfies an ordinary central limit theorem.

Since $\hat{\theta}_G$ is $n^{\frac{1}{2}}$ -consistent, the theorem now follows from Theorem 5.45 of van der Vaart (1998) provided (26) holds and

$$\sup_{\|\theta - \theta_0\| \leq \delta_n} n^{\frac{1}{2}} \{ \bar{S}_W(\theta) - \bar{S}_W(\theta_0) - \mathcal{S}_W(\theta) \} = o_p(1 + n^{\frac{1}{2}} \|\theta - \theta_0\|) \quad (40)$$

for any sequence $\delta_n \rightarrow 0$. Arguing exactly as in the proof of Theorem 1, (A1)-(A10) are sufficient to ensure that both $\bar{S}_W(\theta) - \bar{S}_W(\theta_0)$ and $\mathcal{S}_G(\theta)$ equal

$$-(\theta - \theta_0)' \int_0^{\bar{\tau}} w(u|\theta) v(u|\theta) s^{(0)}(u|\theta) Q(u) \lambda_0(u) du \quad (41)$$

to $o_p(n^{-1/2} + \|\theta - \theta_0\|)$ and $o(\|\theta - \theta_0\|)$, respectively. This establishes (40) and also (26), which follows directly from (41). \square

PROOF OF THEOREM 3:

Conditionally on the data, it is easy to show that

$$\lim_{B \rightarrow \infty} \frac{1}{B} \mathcal{Z}' \mathcal{Z} = \begin{pmatrix} 1 & 0' \\ 0 & n \hat{\Gamma}_G \end{pmatrix} \quad \text{and} \quad \lim_{B \rightarrow \infty} \frac{1}{B} \mathcal{Z}' \mathcal{Y} = \begin{pmatrix} 0' \\ n \hat{\Gamma}_G D_W \end{pmatrix} + \hat{\Gamma}_G o_p(n^{1/2}).$$

Thus, $\hat{D}_W = D_W + o_p(n^{-1/2})$, proving the first result. The second result is a simple consequence of the first and the fact that $\hat{\Sigma}_W \xrightarrow{p} \Sigma_W$ under (A1)-(A10). \square

PROOF OF THEOREM 4:

For any θ , we may write

$$\hat{\Lambda}_0(t|\theta) - \Lambda_0(t) = [\hat{\Lambda}_0(t|\theta) - \hat{\Lambda}_0(t|\theta_0)] + [\hat{\Lambda}_0(t|\theta_0) - \Lambda_0(t)].$$

Consider first $\hat{\Lambda}_0(t|\theta) - \hat{\Lambda}_0(t|\theta_0)$, to which we add $\Lambda_0(t) = \int_0^t [\mathbb{S}^{(0)}(u|\theta)/\mathbb{S}^{(0)}(u|\theta)] \lambda_0(u) du$ and subtract $\Lambda_0(t) = \int_0^t [\mathbb{S}^{(0)}(u|\theta_0)/\mathbb{S}^{(0)}(u|\theta_0)] \lambda_0(u) du$. A suitable rearrangement of terms, several applications of Corollary 1, and arguments similar to those used in the proofs of Theorems 1 and 2 lead to the representation

$$\hat{\Lambda}_0(t|\theta) - \hat{\Lambda}_0(t|\theta_0) \equiv \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{Y_i(u|\theta)}{s^{(0)}(u|\theta)} [\lambda_0(u h_i(\theta)) h_i(\theta) - \lambda_0(u)] du + o_p(n^{-1/2}) \quad (42)$$

where $h_i(\theta) = \exp\{(\theta_0 - \theta)' Z_i\}$ and the $o_p(n^{-1/2})$ term is uniform in t and $\theta \in \mathcal{N}(\theta_0)$. Supposing $\|\theta - \theta_0\| \leq \delta_n$ for any $\delta_n \rightarrow 0$ and proceeding similarly to the proofs of Theorems 1 and 2, the first term on the right-hand side of (42) can be written

$$- \left[\int_0^t \frac{s^{(1)}(u|\theta_0)}{s^{(0)}(u|\theta_0)} Q(u) \lambda_0(u) du \right]' (\theta - \theta_0) + o_p(\|\theta - \theta_0\|),$$

where the $o_p(\|\theta - \theta_0\|)$ term is uniform in t by (A1). Thus, for $\|\theta - \theta_0\| \leq n^{-1/2}M$,

$$\widehat{\Lambda}_0(t|\theta) - \widehat{\Lambda}_0(t|\theta_0) = -A'_Q(t)(\theta - \theta_0) + o_p(n^{-1/2})$$

where $A_Q(t)$ is as specified in the theorem and the $o_p(n^{-1/2})$ term is uniform in t and $\theta \in \mathcal{N}(\theta_0)$. This establishes (28). To establish the first result, note that Corollary 1 implies

$$n^{\frac{1}{2}} \left(\widehat{\Lambda}_0(t|\theta_0) - \Lambda_0(t) \right) = n^{-1/2} \sum_{i=1}^n \int_0^t \frac{M_i(du|\theta_0)}{s^{(0)}(u|\theta)} + o_p(1), \quad (43)$$

the error being uniform for $t \in [0, \bar{\tau}]$. Hence, uniformly for $t \in [0, \bar{\tau}]$, (28), (43), and the fact that $\widehat{\theta}_W - \theta_0 = -[D_W^{-1}]' \bar{S}_W(\theta_0) + o_p(n^{-1/2})$ imply that $n^{\frac{1}{2}} \left(\widehat{\Lambda}_0(t|\widehat{\theta}_W) - \Lambda_0(t) \right)$ is equivalent to

$$n^{-1/2} \sum_{i=1}^n \left[\int_0^t \frac{M_i(du|\theta_0)}{s^{(0)}(u|\theta_0)} + [A_Q(t)D_W^{-1}]' \int_0^{\bar{\tau}} w(u|\theta_0) \left\{ Z_i - \frac{s^{(1)}(u|\theta_0)}{s^{(0)}(u|\theta_0)} \right\} M_i(du|\theta_0) \right] + o_p(1).$$

Write the first term as $n^{-1/2} \sum_{i=1}^n \xi_i(t|\theta_0)$. Under the assumptions of the theorem, $\xi(\cdot|\theta_0)$ is cadlag in t and of bounded variation for $t \in [0, \bar{\tau}]$. Hence, it is \mathbb{P}_0 -Donsker and therefore $n^{-1/2} \sum_{i=1}^n \xi_i(t|\theta_0)$ converges weakly to a Gaussian process with covariance $E[\xi_1(s|\theta_0)\xi_1(t|\theta_0)]$ for any $t, s \geq 0$. Tedious but straightforward computations using Lemma 3 yield the covariance term; see also the proof of Proposition 2. The remaining details are omitted. \square

PROOF OF THEOREM 5:

For $j = 0, 1, 2$, each of (16) and (32) is a monotone function of $t \in \mathbb{R}^+$. Since Z_i is uniformly bounded, there exists $M_j < \infty$ depending only on j such that $\bar{N}^{(j)}(t|\theta) \leq M_j N_i$ and $\mathbb{S}^{(j)}(t|\theta) \leq M_j(N_i + 1)$. Under (A5), the expectation of $(N_i + 1)^2$ is finite under \mathbb{P}_0 . Hence each is Donsker by Example 2.10.27 of van der Vaart and Wellner (1996). \square

PROOF OF THEOREM 6:

Let the observed data on subject i be $\mathfrak{X}_i = (\{(N_i^\dagger(u), Y_i^\dagger(u)), u \geq 0\}, Z_i)$; our sample then constitutes n iid copies of a generic element \mathfrak{X} defined on the relevant sample space \mathcal{X} . The processes in (16) and (32) are each constructed via a Z -weighted \mathfrak{X} -dependent sum of indicator functions, where each indicator function is parameterized by (t, θ) in a similar manner. The argument required for establishing the Donsker property is therefore the same for both types of processes, and it suffices to establish that (32) is \mathbb{P}_0 -Donsker. For a generic element $\mathfrak{X} \in \mathcal{X}$, define

$$f(\mathfrak{X}, t, \theta) = \sum_{j=1}^{k(\mathfrak{X})} I\{g_j(\mathfrak{X}, t, \theta) \leq 0\}, \quad (44)$$

where $k(\mathfrak{X}_i) = N_i^\dagger(\tau_i-) = N_i$ and $g_j(\mathfrak{X}_i, t, \theta) = T_{ij}e^{\theta'Z_i} - t$ for $j \leq N_i$. Clearly, $f(\mathfrak{X}_i, t, \theta) \equiv N_i(t|\theta)$ and thus $\bar{N}^{(0)}(t|\theta) \equiv n^{-1} \sum_{i=1}^n f(\mathfrak{X}_i, t, \theta)$. Define

$$\mathcal{F} = \left\{ f(\cdot, t, \theta) = \sum_{j=1}^{k(\cdot)} I\{g_j(\cdot, t, \theta) \leq 0\}, t \in \mathbb{R}^+, \theta \in \Theta \right\}; \quad (45)$$

Considered as a function on $\mathbb{R}^+ \times \Theta$, $f(\mathfrak{X}_i, t, \theta)$ is evidently an element of \mathcal{F} . Now, by (A5), $Z_1 \dots Z_n$ are uniformly bounded in norm. If (45) is a \mathbb{P}_0 -Donsker class with integrable envelope, the class of functions formed by computing $Zf(\mathfrak{X}, t, \theta)$ for $f \in \mathcal{F}$ is also \mathbb{P}_0 -Donsker (van der Vaart and Wellner, 1996, Ex. 2.10.10). Hence it suffices to establish the Donsker property for $j = 0$, i.e., for \mathcal{F} . The elements (44) of \mathcal{F} are constructed by summing simpler functions of \mathfrak{X} and (t, θ) . The proof therefore proceeds by showing the Donsker property holds for these simpler functions. Theorem 2.10.24 of van der Vaart and Wellner (1996) and other results on preservation of the Donsker property are then used to establish the result for \mathcal{F} . The remaining details are available from the author upon request. \square

REFERENCES

- Andersen, P.K. & Gill, R.D. (1982). Cox's regression model for counting processes: a large sample study. *Annals of Statistics* **10**, 1100-1120.
- Andersen, P.K., Borgan, O., Gill, R.D., & Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.
- Bickel, P.J., Klaassen, C.A.J., Ritov, Y., & Wellner, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.
- Borgan, O. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scandinavian Journal of Statistics* **11**, 1-16.
- Byar, D. (1980). The Veterans Administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo, pyridoxine, and topical thiotepa. In *Bladder Tumors and Other Topics in Urological Oncology*, Ed. M. Pavone-Macaluso, P.H. Smith and F. Fedsmyn, pp. 363-370. New York: Plenum.
- Chang, I.S., & Hsiung, C.A. (1994). Information and asymptotic efficiency in some generalized proportional hazards models for counting processes. *Annals of Statistics* **22**, 1275-1298.
- Chang, S., & Wang, M.C. (1999). Conditional regression analysis for recurrence time data. *Journal of the American Statistical Association* **94**, 1221-1230.
- Cox, D. (1972a). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Ser. B* **34**, 187-220.
- Cox, D.R. (1972b). The statistical analysis of dependencies in point processes. In *Stochastic point processes: statistical analysis, theory, and applications*, Ed. P.A.W. Lewis, pp. 55-66. New York: Wiley.
- Cox, D.R. & Solomon, P.J. (2003). *Components of Variance*. Boca Raton: Chapman & Hall.
- Fleming, T.R. & Harrington, D.P. (1992). *Counting Processes and Survival Analysis*. New York: Wiley.
- Fygenson, M., & Ritov, Y. (1994). Monotone estimating equations for censored data. *Annals of Statistics* **22**, 732-746.
- Gill, R.D. (1980). Nonparametric estimation based on censored observations of a Markov renewal process. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **53**, 97-116.
- Gill, R.D. (1981). Testing with replacement and the product-limit estimator. *Annals of Statistics* **9**, 853-860.
- Huang, Y. (2002). Calibration regression of censored lifetime medical cost. *Journal of the American Statistical Association* **97**, 318-327.
- Jin, Z., Lin, D.Y., Wei, L.J., & Ying Z. (1993). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341-353.
- Kalbfleisch, J.D. & Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd ed.. New York: Wiley.
- Lawless, J. & Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37**, 158-285.
- Lin, D.Y., Wei, L.J., & Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85**, 605-618.
- Lin, D.Y., Wei, L.J., Yang, I., & Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society, Ser. B* **62**, 711-730.

- Lin, D.Y., Wei, L.J., & Ying, Z. (2001). Semiparametric transformation models for point processes. *Journal of the American Statistical Association* **96**, 620-628.
- Newey, W.K. & McFadden, D.L. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, Vol. IV, Ed. R.F. Engle and D.L. McFadden, pp. 2111-2245. Amsterdam: Elsevier.
- Oakes, D. (1992). Frailty models for multiple event times. In *Survival Analysis: State of the Art*, Ed. J. Klein and P. Goel, pp. 371-379. Dordrecht: Kluwer.
- Oakes, D., & Cui, L. (1994). On semiparametric inference for modulated renewal processes. *Biometrika* **81**, 83-90.
- Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Annals of Statistics* **26**, 183-214.
- Peña, E.A., Strawderman, R.L., & Hollander, M. (2000). A weak convergence result relevant in recurrent and renewal models. In *Recent Advances in Reliability Theory: Methodology, Practice and Inference*, Ed. N. Limnios and M. Nikulin, pp. 493-514. Boston: Birkhauser.
- Peña, E.A., Strawderman, R.L., & Hollander, M. (2001). Nonparametric estimation with recurrent event data. *Journal of the American Statistical Association* **96**, 1299-1315.
- Pepe, M.S., & Cai J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time-dependent covariates. *Journal of the American Statistical Association* **88**, 811-820.
- Prentice, R.L., Williams, B., & Petersen, A. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68**, 373-379.
- Reid, N. (1994). A conversation with Sir David Cox. *Statistical Science* **9**, 439-455.
- Resnick, S. (1994). *Adventures in Stochastic Processes*. Birkhauser: Boston.
- Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Annals of Statistics* **18**, 303-328.
- Ritov, Y. (1991). Estimating functions in semi-parametric models. In *Estimating Functions*, Ed. V. Godambe, p. 319-336. Oxford: Clarendon Press.
- Robins, J., & Tsiatis, A.A. (1992). Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika* **79**, 311-319.
- Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York: Wiley.
- Tsiatis, A.A. (1990). Estimating regression parameters using linear rank tests for censored data. *Annals of Statistics* **18**, 354-372.
- van de Geer, S. (2000). *Empirical Processes in M-Estimation*. Cambridge: Cambridge University Press.
- van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.
- van der Vaart, A.W. & Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.
- Wang, M.C. & Chang, S. (1999). Nonparametric estimation of recurrent survival function. *Journal of the American Statistical Association* **94**, 146-153.
- Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *Annals of Statistics* **21**, 76-99.